

EXPRESSED SEQUENCES OF *ARABIDOPSIS THALIANA*5 *CROSS-REFERENCE TO RELATED APPLICATION*

This application claims the benefit of U.S. Provisional Application 60/178,472
Filed January 27, 2000.

FIELD OF INVENTION

10 The invention is in the field of polynucleotide sequences of a plant, particularly
sequences expressed in *arabidopsis thaliana*.

BACKGROUND OF THE INVENTION

Plants and plant products have vast commercial importance in a wide variety of
areas including food crops for human and animal consumption, flavor enhancers for
15 food, and production of specialty chemicals for use in products such as medicaments
and fragrances. In considering food crops for humans and livestock, genes such as
those involved in a plant's resistance to insects, plant viruses, and fungi; genes involved
in pollination; and genes whose products enhance the nutritional value of the food, are
of major importance. A number of such genes have been described, see, for example,
20 McCaskill and Croteau (1999) *Nature Biotechnol.* 17:31-36.

Despite recent advances in methods for identification, cloning, and
characterization of genes, much remains to be learned about plant physiology in
general, including how plants produce many of the above-mentioned products;
mechanisms for resistance to herbicides, insects, plant viruses, fungi; elucidation of
25 genes involved in specific biosynthetic pathways; and genes involved in environmental
tolerance, e.g., salt tolerance, drought tolerance, or tolerance to anaerobic conditions.

Arabidopsis thaliana is a model system for genetic, molecular and biochemical
studies of higher plants. Features of this plant that make it a model system for genetic
and molecular biology research include a small genome size, organized into five
30 chromosomes and containing an estimated 20,000 genes, a rapid life cycle, prolific seed
production and, since it is small, it can easily be cultivation in limited space. *A. thaliana*

is a member of the mustard family (*Brassicaceae*) with a broad natural distribution throughout Europe, Asia, and North America. Many different ecotypes have been collected from natural populations and are available for experimental analysis. The entire life cycle, including seed germination, formation of a rosette plant, bolting of the main stem, flowering, and maturation of the first seeds, is completed in 6 weeks. A large number of mutant lines are available that affect nearly all aspects of its growth. These features greatly facilitate the isolation of fundamentally interesting and potentially important genes for agronomic development

Most gene products from higher plants exhibit adequate sequence similarity to deduced amino acid sequences of other plant genes to permit assignment of probable gene function, if it is known, in any higher plant. It is likely that there will be very few protein-encoding angiosperm genes that do not have orthologs or paralogs in *Arabidopsis*. The developmental diversity of higher plants may be largely due to changes in the cis-regulatory sequences of transcriptional regulators and not in coding sequences.

Many advances reported over the past few years offer clear evidence that this plant is not only a very important model species for basic research, but also extremely valuable for applied plant scientists and plant breeders. Knowledge gained from *Arabidopsis* can be used directly to develop desired traits in plants of other species.

Relevant Literature

Cold Spring Harbor Monograph 27 (1994) E.M. Meyerowitz and C.R. Somerville, eds. (CSH Laboratory Press). Annual Plant Reviews, Vol. 1: *Arabidopsis* (1998) M. Anderson and J.A. Roberts, eds. (CRC Press). Methods in Molecular Biology: *Arabidopsis* Protocols, Vol. 82 (1997) J.M. Martinez-Zapater and J. Salinas, eds. (CRC Press).

Mayer *et al* (1999) Nature **402**(6763):769-77; "Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*". Lin *et al.* (1999) **402**(6763):761-8, "Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*". Meinke *et al.* (1998) Science **282**:662-682, "*Arabidopsis thaliana*: a model plant for genome analysis". Somerville and Somerville (1999) Science **285**:380-383, "Plant functional

genomics". Mozo *et al.* (1999) Nat. Genet. **22**:271-275, "A complete BAC-based physical map of the *Arabidopsis thaliana* genome".

SUMMARY OF THE INVENTION

5 Novel nucleic acid sequences of *Arabidopsis thaliana*, their encoded polypeptides and variants thereof, genes corresponding to these nucleic acids, and proteins expressed by the genes, are provided.

The invention also provides diagnostic, prophylactic and therapeutic agents employing such novel nucleic acids, their corresponding genes or gene products,
10 including expression constructs, probes, antisense constructs, and the like. The genetic sequences may also be used for the genetic manipulation of plant cells, particularly dicotyledonous plants. The encoded gene products and modified organisms are useful for introducing or improving disease resistance and stress tolerance into plants; screening of biologically active agents, *e.g.* fungicides, *etc.*; for elucidating biochemical
15 pathways; and the like.

In one embodiment of the invention, a nucleic acid is provided that comprises a start codon; an optional intervening sequence; a coding sequence capable of hybridizing under stringent conditions as set forth in SEQ ID NO:1 to 999; and an optional terminal sequence, wherein at least one of said optional sequences is present. Such a nucleic
20 acid may correspond to naturally occurring *Arabidopsis* expressed sequences.

DETAILED DESCRIPTION OF THE INVENTION

Novel nucleic acid sequences from *Arabidopsis thaliana*, their encoded polypeptides and variants thereof, genes corresponding to these nucleic acids and
25 proteins expressed by the genes are provided. The invention also provides agents employing such novel nucleic acids, their corresponding genes or gene products, including expression constructs, probes, antisense constructs, and the like. The nucleotide sequences are provided in the attached SEQLIST.

Sequences include, but are not limited to, sequences that encode resistance
30 proteins; sequences that encode tolerance factors; sequences encoding proteins or other factors that are involved, directly or indirectly in biochemical pathways such as

metabolic or biosynthetic pathways, sequences involved in signal transduction, sequences involved in the regulation of gene expression, structural genes, and the like.

Biosynthetic pathways of interest include, but are not limited to, biosynthetic pathways whose product (which may be an end product or an intermediate) is of commercial, nutritional, or medicinal value.

The sequences may be used in screening assays of various plant strains to determine the strains that are best capable of withstanding a particular disease or environmental stress. Sequences encoding activators and resistance proteins may be introduced into plants that are deficient in these sequences. Alternatively, the sequences may be introduced under the control of promoters that are convenient for induction of expression. The protein products may be used in screening programs for insecticides, fungicides and antibiotics to determine agents that mimic or enhance the resistance proteins. Such agents may be used in improved methods of treating crops to prevent or treat disease. The protein products may also be used in screening programs to identify agents which mimic or enhance the action of tolerance factors. Such agents may be used in improved methods of treating crops to enhance their tolerance to environmental stresses.

Still other embodiments of the invention provide methods for enhancing or inhibiting production of a biosynthetic product in a plant by introducing a nucleic acid of the invention into a plant cell, where the nucleic acid comprises sequences encoding a factor which is involved, directly or indirectly in a biosynthetic pathway whose products are of commercial, nutritional, or medicinal value include any factor, usually a protein or peptide, which regulates such a biosynthetic pathway; which is an intermediate in such a biosynthetic pathway; or which in itself is a product that increases the nutritional value of a food product; or which is a medicinal product; or which is any product of commercial value.

Transgenic plants containing the antisense nucleic acids of the invention are useful for identifying other mediators that may induce expression of proteins of interest; for establishing the extent to which any specific insect and/or pathogen is responsible for damage of a particular plant; for identifying other mediators that may enhance or induce tolerance to environmental stress; for identifying factors involved in biosynthetic

pathways of nutritional, commercial, or medicinal value; or for identifying products of nutritional, commercial, or medicinal value.

5 In still other embodiments, the invention provides transgenic plants constructed by introducing a subject nucleic acid of the invention into a plant cell, and growing the cell into a callus and then into a plant; or, alternatively by breeding a transgenic plant from the subject process with a second plant to form an F1 or higher hybrid. The subject transgenic plants and progeny are used as crops for their enhanced disease resistance, enhanced traits of interest, for example size or flavor of fruit, length of growth cycle, *etc.*, or for screening programs, *e.g.* to determine more effective insecticides, *etc.*
10 used as crops which exhibit enhanced tolerance environmental stress; or used to produce a factor.

Those skilled in the art will recognize the agricultural advantages inherent in plants constructed to have either increased or decreased expression of resistance proteins; or increased or decreased tolerance to environmental factors; or which
15 produce or over-produce one or more factors involved in a biosynthetic pathway whose product is of commercial, nutritional, or medicinal value. For example, such plants may have increased resistance to attack by predators, insects, pathogens, microorganisms, herbivores, mechanical damage and the like; may be more tolerant to environmental stress, *e.g.* may be better able to withstand drought conditions, freezing, and the like;
20 or may produce a product not normally made in the plant, or may produce a product in higher than normal amounts, where the product has commercial, nutritional, or medicinal value. Plants which may be useful include dicotyledons and monocotyledons.

Representative examples of plants in which the provided sequences may be useful include tomato, potato, tobacco, cotton, soybean, alfalfa, rape, and the like.
25 Monocotyledons, more particularly grasses (*Poaceae* family) of interest, include, without limitation, *Avena sativa* (oat); *Avena strigosa* (black oat); *Elymus* (wild rye); *Hordeum* *sp.* including *Hordeum vulgare* (barley); *Oryza* *sp.*, including *Oryza glaberrima* (African rice); *Oryza longistaminata* (long-staminate rice); *Pennisetum americanum* (pearl millet); *Sorghum* *sp.* (sorghum); *Triticum* *sp.*, including *Triticum aestivum* (common wheat);
30 *Triticum durum* (durum wheat); *Zea mays* (corn); *etc.*

NUCLEIC ACID COMPOSITIONS

The following detailed description describes the nucleic acid compositions encompassed by the invention, methods for obtaining cDNA or genomic DNA encoding a full-length gene product, expression of these nucleic acids and genes; identification
5 of structural motifs of the nucleic acids and genes; identification of the function of a gene product encoded by a gene corresponding to a nucleic acid of the invention; use of the provided nucleic acids as probes, in mapping, and in diagnosis; use of the corresponding polypeptides and other gene products to raise antibodies; use of the nucleic acids in genetic modification of plant and other species; and use of the nucleic
10 acids, their encoded gene products, and modified organisms, for screening and diagnostic purposes.

The scope of the invention with respect to nucleic acid compositions includes, but is not necessarily limited to, nucleic acids having a sequence set forth in any one of SEQ ID NOS:1-999; nucleic acids that hybridize the provided sequences under stringent
15 conditions; genes corresponding to the provided nucleic acids; variants of the provided nucleic acids and their corresponding genes, particularly those variants that retain a biological activity of the encoded gene product.

In one embodiment, the sequences of the invention provide a polypeptide coding sequence. The polypeptide coding sequence may correspond to a naturally expressed
20 mRNA in Arabidopsis or other species, or may encode a fusion protein between one of the provided sequences and an exogenous protein coding sequence. The coding sequence is characterized by an ATG start codon, a lack of stop codons in-frame with the ATG, and a termination codon, that is, a continuous open frame is provided between the start and the stop codon. The sequence contained between the start and the stop
25 codon will comprise a sequence capable of hybridizing under stringent conditions to a sequence set for in SEQ ID NO:1-999, and may comprise the sequence set forth in the Seqlist.

Other nucleic acid compositions contemplated by and within the scope of the present invention will be readily apparent to one of ordinary skill in the art when provided
30 with the disclosure here.

The invention features nucleic acids that are derived from *Arabidopsis thaliana*. Novel nucleic acid compositions of the invention of particular interest comprise a sequence set forth in any one of SEQ ID NOS:1-999 or an identifying sequence thereof.

An "identifying sequence" is a contiguous sequence of residues at least about 10 nt to about 20 nt in length, usually at least about 50 nt to about 100 nt in length, that uniquely identifies a nucleic acid sequence, e.g., exhibits less than 90%, usually less than about 80% to about 85% sequence identity to any contiguous nucleotide sequence of more than about 20 nt. Thus, the subject novel nucleic acid compositions include full length cDNAs or mRNAs that encompass an identifying sequence of contiguous nucleotides from any one of SEQ ID NOS:1-999.

The nucleic acids of the invention also include nucleic acids having sequence similarity or sequence identity. Nucleic acids having sequence similarity are detected by hybridization under low stringency conditions, for example, at 50°C and 10XSSC (0.9 M NaCl/0.09 M sodium citrate) and remain bound when subjected to washing at 55°C in 1XSSC. Sequence identity can be determined by hybridization under stringent conditions, for example, at 50°C or higher and 0.1XSSC (9 mM NaCl/0.9 mM sodium citrate). Hybridization methods and conditions are well known in the art, see U.S. Patent No. 5,707,829. Nucleic acids that are substantially identical to the provided nucleic acid sequences, e.g. allelic variants, genetically altered versions of the gene, etc., bind to the provided nucleic acid sequences (SEQ ID NOS:1-999) under stringent hybridization conditions. By using probes, particularly labeled probes of DNA sequences, one can isolate homologous or related genes. The source of homologous genes can be any species, particularly grasses as previously described.

Preferably, hybridization is performed using at least 15 contiguous nucleotides of at least one of SEQ ID NOS:1-999. The probe will preferentially hybridize with a nucleic acid or mRNA comprising the complementary sequence, allowing the identification and retrieval of the nucleic acids of the biological material that uniquely hybridize to the selected probe. Probes of more than 15 nucleotides can be used, e.g. probes of from about 18 nucleotides up to the entire length of the provided nucleic acid sequences, but 15 nucleotides generally represents sufficient sequence for unique identification.

The nucleic acids of the invention also include naturally occurring variants of the nucleotide sequences, e.g. degenerate variants, allelic variants, etc. Variants of the nucleic acids of the invention are identified by hybridization of putative variants with nucleotide sequences disclosed herein, preferably by hybridization under stringent conditions. For example, by using appropriate wash conditions, variants of the nucleic acids of the invention can be identified where the allelic variant exhibits at most about 25-30% base pair mismatches relative to the selected nucleic acid probe. In general, allelic variants contain 5-25% base pair mismatches, and can contain as little as even 2-5%, or 1-2% base pair mismatches, as well as a single base-pair mismatch.

The invention also encompasses homologs corresponding to the nucleic acids of SEQ ID NOS:1-999, where the source of homologous genes can be any related species, usually within the same genus or group. Homologs have substantial sequence similarity, e.g. at least 75% sequence identity, usually at least 90%, more usually at least 95% between nucleotide sequences. Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, etc. A reference sequence will usually be at least about 18 contiguous nt long, more usually at least about 30 nt long, and may extend to the complete sequence that is being compared. Algorithms for sequence analysis are known in the art, such as BLAST, described in Altschul et al., J. Mol. Biol. (1990) 215:403-10.

In general, variants of the invention have a sequence identity greater than at least about 65%, preferably at least about 75%, more preferably at least about 85%, and can be greater than at least about 90% or more as determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular). For the purposes of this invention, a preferred method of calculating percent identity is the Smith-Waterman algorithm, using the following. Global DNA sequence identity must be greater than 65% as determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular) using an affine gap search with the following search parameters: gap open penalty, 12; and gap extension penalty,

1.

The subject nucleic acids can be cDNAs or genomic DNAs, as well as fragments thereof, particularly fragments that encode a biologically active gene product and/or are useful in the methods disclosed herein. The term "cDNA" as used herein is intended to include all nucleic acids that share the arrangement of sequence elements found in native mature mRNA species, where sequence elements are exons and 3' and 5' non-coding regions. Normally mRNA species have contiguous exons, with the introns, when present, being removed by nuclear RNA splicing, to create a continuous open reading frame encoding a polypeptide of the invention.

A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It can further include the 3' and 5' untranslated regions found in the mature mRNA. It can further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, etc., including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' and 3' end of the transcribed region. The genomic DNA can be isolated as a fragment of 100 kb or smaller; and substantially free of flanking chromosomal sequence. The genomic DNA flanking the coding region, either 3' and 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for expression.

The nucleic acid compositions of the subject invention can encode all or a part of the subject expressed polypeptides. Double or single stranded fragments can be obtained from the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, etc. Isolated nucleic acids and nucleic acid fragments of the invention comprise at least about 15 up to about 100 contiguous nucleotides, or up to the complete sequence provided in SEQ ID NOS:1-999. For the most part, fragments will be of at least 15 nt, usually at least 18 nt or 25 nt, and up to at least about 50 contiguous nt in length or more.

Probes specific to the nucleic acids of the invention can be generated using the nucleic acid sequences disclosed in SEQ ID NOS:1-999 and the fragments as described above. The probes can be synthesized chemically or can be generated from longer nucleic acids using restriction enzymes. The probes can be labeled, for example, with

a radioactive, biotinylated, or fluorescent tag. Preferably, probes are designed based upon an identifying sequence of a nucleic acid of one of SEQ ID NOS:1-999. More preferably, probes are designed based on a contiguous sequence of one of the subject nucleic acids that remain unmasked following application of a masking program for
5 masking low complexity (e.g., XBLAST) to the sequence., *i.e.* one would select an unmasked region, as indicated by the nucleic acids outside the poly-n stretches of the masked sequence produced by the masking program.

The nucleic acids of the subject invention are isolated and obtained in substantial purity, generally as other than an intact chromosome. Usually, the nucleic acids, either
10 as DNA or RNA, will be obtained substantially free of other naturally-occurring nucleic acid sequences, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", e.g., flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

The nucleic acids of the invention can be provided as a linear molecule or within
15 a circular molecule. They can be provided within autonomously replicating molecules (vectors) or within molecules without replication sequences. They can be regulated by their own or by other regulatory sequences, as is known in the art. The nucleic acids of the invention can be introduced into suitable host cells using a variety of techniques which are available in the art, such as transferrin polycation-mediated DNA transfer,
20 transfection with naked or encapsulated nucleic acids, liposome-mediated DNA transfer, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, gene gun, calcium phosphate-mediated transfection, and the like.

The subject nucleic acid compositions can be used to, for example, produce polypeptides, as probes for the detection of mRNA of the invention in biological
25 samples, e.g. extracts of cells, to generate additional copies of the nucleic acids, to generate ribozymes or antisense oligonucleotides, and as single stranded DNA probes or as triple-strand forming oligonucleotides. The probes described herein can be used to, for example, determine the presence or absence of the nucleic acid sequences as shown in SEQ ID NOS:1-999 or variants thereof in a sample. These and other uses are
30 described in more detail below.

USE OF NUCLEIC ACIDS AS CODING SEQUENCES

Naturally occurring Arabidopsis polypeptides or fragments thereof are encoded by the provided nucleic acids. Methods are known in the art to determine whether the complete native protein is encoded by a candidate nucleic acid sequence. Where the provided sequence encodes a fragment of a polypeptide, methods known in the art may be used to determine the remaining sequence. These approaches may utilize a bioinformatics approach, a cloning approach, extension of mRNA species, *etc.*

Substantial genomic sequence is available for Arabidopsis, and may be exploited for determining the complete coding sequence corresponding to the provided sequences. The region of the chromosome to which a given sequence is located may be determined by hybridization or by database searching. The genomic sequence is then searched upstream and downstream for the presence of intron/exon boundaries, and for motifs characteristic of transcriptional start and stop sequences, for example by using Genscan (Burge and Karlin (1997) J. Mol. Biol. **268**:78-94); or GRAIL (Uberbacher and Mural (1991) P.N.A.S. **88**:11261-1265).

Alternatively, nucleic acid having a sequence of one of SEQ ID NOS:1-999, or an identifying fragment thereof, is used as a hybridization probe to complementary molecules in a cDNA library using probe design methods, cloning methods, and clone selection techniques as known in the art. Libraries of cDNA are made from selected cells. The cells may be those of *A. thaliana*, or of related species. In some cases it will be desirable to select cells from a particular stage, *e.g.* seeds, leaves, infected cells, *etc.*

Techniques for producing and probing nucleic acid sequence libraries are described, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY; and *Current Protocols in Molecular Biology*, (1987 and updates) Ausubel et al., eds. The cDNA can be prepared by using primers based on sequence from SEQ ID NOS:1-999. In one embodiment, the cDNA library can be made from only poly-adenylated mRNA. Thus, poly-T primers can be used to prepare cDNA from the mRNA.

Members of the library that are larger than the provided nucleic acids, and preferably that encompass the complete coding sequence of the native message, are obtained. In order to confirm that the entire cDNA has been obtained, RNA protection

experiments are performed as follows. Hybridization of a full-length cDNA to an mRNA will protect the RNA from RNase degradation. If the cDNA is not full length, then the portions of the mRNA that are not hybridized will be subject to RNase degradation. This is assayed, as is known in the art, by changes in electrophoretic mobility on polyacrylamide gels, or by detection of released monoribonucleotides. Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY. In order to obtain additional sequences 5' to the end of a partial cDNA, 5' RACE (PCR Protocols: A Guide to Methods and Applications, (1990) Academic Press, Inc.) may be performed.

Genomic DNA is isolated using the provided nucleic acids in a manner similar to the isolation of full-length cDNAs. Briefly, the provided nucleic acids, or portions thereof, are used as probes to libraries of genomic DNA. Preferably, the library is obtained from the cell type that was used to generate the nucleic acids of the invention, but this is not essential. Such libraries can be in vectors suitable for carrying large segments of a genome, such as P1 or YAC, as described in detail in Sambrook et al., 9.4-9.30. In order to obtain additional 5' or 3' sequences, chromosome walking is performed, as described in Sambrook *et al.*, such that adjacent and overlapping fragments of genomic DNA are isolated. These are mapped and pieced together, as is known in the art, using restriction digestion enzymes and DNA ligase.

PCR methods may be used to amplify the members of a cDNA library that comprise the desired insert. In this case, the desired insert will contain sequence from the full length cDNA that corresponds to the instant nucleic acids. Such PCR methods include gene trapping and RACE methods. Gene trapping entails inserting a member of a cDNA library into a vector. The vector then is denatured to produce single stranded molecules. Next, a substrate-bound probe, such a biotinylated oligo, is used to trap cDNA inserts of interest. Biotinylated probes can be linked to an avidin-bound solid substrate. PCR methods can be used to amplify the trapped cDNA. To trap sequences corresponding to the full length genes, the labeled probe sequence is based on the nucleic acid sequences of the invention. Random primers or primers specific to the library vector can be used to amplify the trapped cDNA. Such gene trapping techniques are described in Gruber *et al.*, WO 95/04745 and Gruber *et al.*, U.S. Pat. No. 5,500,356.

Kits are commercially available to perform gene trapping experiments from, for example, Life Technologies, Gaithersburg, Maryland, USA.

“Rapid amplification of cDNA ends”, or RACE, is a PCR method of amplifying cDNAs from a number of different RNAs. The cDNAs are ligated to an oligonucleotide linker, and amplified by PCR using two primers. One primer is based on sequence from the instant nucleic acids, for which full length sequence is desired, and a second primer comprises sequence that hybridizes to the oligonucleotide linker to amplify the cDNA.

A description of this methods is reported in WO 97/19110. A common primer may be designed to anneal to an arbitrary adaptor sequence ligated to cDNA ends. When a single gene-specific RACE primer is paired with the common primer, preferential amplification of sequences between the single gene specific primer and the common primer occurs. Commercial cDNA pools modified for use in RACE are available.

Once the full-length cDNA or gene is obtained, DNA encoding variants can be prepared by site-directed mutagenesis, described in detail in Sambrook et al., 15.3-15.63. The choice of codon or nucleotide to be replaced can be based on disclosure herein on optional changes in amino acids to achieve altered protein structure and/or function. As an alternative method to obtaining DNA or RNA from a biological material, nucleic acid comprising nucleotides having the sequence of one or more nucleic acids of the invention can be synthesized.

EXPRESSION OF POLYPEPTIDES

The provided nucleic acid, e.g. a nucleic acid having a sequence of one of SEQ ID NOS:1-999), the corresponding cDNA, the polypeptide coding sequence as described above, or the full-length gene is used to express a partial or complete gene product. Constructs of nucleic acids having sequences of SEQ ID NOS:1-999 can be generated by recombinant methods, synthetically, or in a single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides is described by, e.g. Stemmer *et al.*, Gene (Amsterdam) (1995) 164(1):49-53.

Appropriate nucleic acid constructs are purified using standard recombinant DNA techniques as described in, for example, Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY.

The gene product encoded by a nucleic acid of the invention is expressed in any expression system, including, for example, bacterial, yeast, insect, amphibian and mammalian systems.

The subject nucleic acid molecules are generally propagated by placing the molecule in a vector. Viral and non-viral vectors are used, including plasmids. The choice of plasmid will depend on the type of cell in which propagation is desired and the purpose of propagation. Certain vectors are useful for amplifying and making large amounts of the desired DNA sequence. Other vectors are suitable for expression in cells in culture. Still other vectors are suitable for transfer and expression in cells in a whole organism or person. The choice of appropriate vector is well within the skill of the art. Many such vectors are available commercially.

The nucleic acids set forth in SEQ ID NOS:1-999 or their corresponding full-length nucleic acids are linked to regulatory sequences as appropriate to obtain the desired expression properties. These can include promoters attached either at the 5' end of the sense strand or at the 3' end of the antisense strand, enhancers, terminators, operators, repressors, and inducers. The promoters can be regulated or constitutive.

In some situations it may be desirable to use conditionally active promoters, such as tissue-specific or developmental stage-specific promoters. These are linked to the desired nucleotide sequence using the techniques described above for linkage to vectors. Any techniques known in the art can be used.

When any of the above host cells, or other appropriate host cells or organisms, are used to replicate and/or express the nucleic acids or nucleic acids of the invention, the resulting replicated nucleic acid, RNA, expressed protein or polypeptide, is within the scope of the invention as a product of the host cell or organism. The product is recovered by any appropriate means known in the art.

IDENTIFICATION OF FUNCTIONAL AND STRUCTURAL MOTIFS

Translations of the nucleotide sequence of the provided nucleic acids, cDNAs or full genes can be aligned with individual known sequences. Similarity with individual sequences can be used to determine the activity of the polypeptides encoded by the nucleic acids of the invention. Also, sequences exhibiting similarity with more than one

individual sequence can exhibit activities that are characteristic of either or both individual sequences.

The six possible reading frames may be translated using programs such as GCG pepdata, or GCG Frames (Wisconsin Package Version 10.0, Genetics Computer Group (GCG) , Madison, Wisconsin, USA.). Programs such as ORFFinder (National Center for Biotechnology Information (NCBI) a division of the National Library of Medicine (NLM) at the National Institutes of Health (NIH) <http://www.ncbi.nlm.nih.gov/>) may be used to identify open reading frames (ORFs) in sequences. ORF finder identifies all possible ORFs in a DNA sequence by locating the standard and alternative stop and start codons. Other ORF identification programs include Genie (Kulp *et al.* (1996).

A generalized Hidden Markov Model may be used for the recognition of genes in DNA. (ISMB-96, St. Louis, MO, AAAI/MIT Press; Reese *et al.* (1997), "Improved splice site detection in Genie". Proceedings of the First Annual International Conference on Computational Molecular Biology RECOMB 1997, Santa Fe, NM, ACM Press, New York., P. 34.); BESTORF --Prediction of potential coding fragment in human or plant EST/mRNA sequence data using Markov Chain Models; and FGENEP -- Multiple genes structure prediction in plant genomic DNA (Solovyev *et al.* (1995) Identification of human gene structure using linear discriminant functions and dynamic programming. In Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology eds. Rawling *et al.* Cambridge, England, AAAI Press, 367-375.; Solovyev *et al.* (1994) Nucl. Acids Res. **22**(24):5156-5163; Solovyev *et al.*, The prediction of human exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames, in: The Second International conference on Intelligent systems for Molecular Biology (eds. Altman *et al.*), AAAI Press, Menlo Park, CA (1994, 354-362) Solovyev and Lawrence, Prediction of human gene structure using dynamic programming and oligonucleotide composition, In: Abstracts of the 4th annual Keck symposium. Pittsburgh, 47, 1993; Burge and Karlin (1997) J. Mol. Biol. **268**:78-94; Kulp *et al.* (1996) Proc. Conf. on Intelligent Systems in Molecular Biology '96, 134-142).

The full length sequences and fragments of the nucleic acid sequences of the nearest neighbors can be used as probes and primers to identify and isolate the full length sequence corresponding to provided nucleic acids. Typically, a selected nucleic

acid is translated in all six frames to determine the best alignment with the individual sequences. These amino acid sequences are referred to, generally, as query sequences, which are aligned with the individual sequences. Suitable databases include Genbank, EMBL, and DNA Database of Japan (DDBJ).

5 Query and individual sequences can be aligned using the methods and computer programs described above, and include BLAST, available by ftp at <ftp://ncbi.nlm.nih.gov/>.

 Gapped BLAST and PSI-BLAST are useful search tools provided by NCBI. (version 2.0) (Altschul *et al.*, 1997). Position-Specific Iterated BLAST (PSI-BLAST)
10 provides an automated, easy-to-use version of a "profile" search, which is a sensitive way to look for sequence homologues. The program first performs a gapped BLAST database search. The PSI-BLAST program uses the information from any significant alignments returned to construct a position-specific score matrix, which replaces the query sequence for the next round of database searching. PSI-BLAST may be iterated
15 until no new significant alignments are found. The Gapped BLAST algorithm allows gaps (deletions and insertions) to be introduced into the alignments that are returned. Allowing gaps means that similar regions are not broken into several segments. The scoring of these gapped alignments tends to reflect biological relationships more closely. The Smith-Waterman is another algorithm that produces local or global gapped
20 sequence alignments, see Meth. Mol. Biol. (1997) 70: 173-187. Also, the GAP program using the Needleman and Wunsch global alignment method can be utilized for sequence alignments.

 Results of individual and query sequence alignments can be divided into three categories, high similarity, weak similarity, and no similarity. Individual alignment results
25 ranging from high similarity to weak similarity provide a basis for determining polypeptide activity and/or structure. Parameters for categorizing individual results include: percentage of the alignment region length where the strongest alignment is found, percent sequence identity, and e value.

 The percentage of the alignment region length is calculated by counting the
30 number of residues of the individual sequence found in the region of strongest alignment, *e.g.* contiguous region of the individual sequence that contains the greatest

number of residues that are identical to the residues of the corresponding region of the aligned query sequence. This number is divided by the total residue length of the query sequence to calculate a percentage. For example, a query sequence of 20 amino acid residues might be aligned with a 20 amino acid region of an individual sequence. The individual sequence might be identical to amino acid residues 5, 9-15, and 17-19 of the query sequence. The region of strongest alignment is thus the region stretching from residue 9-19, an 11 amino acid stretch. The percentage of the alignment region length is: 11 (length of the region of strongest alignment) divided by (query sequence length) 20 or 55%.

Percent sequence identity is calculated by counting the number of amino acid matches between the query and individual sequence and dividing total number of matches by the number of residues of the individual sequences found in the region of strongest alignment. Thus, the percent identity in the example above would be 10 matches divided by 11 amino acids, or approximately, 90.9%

E value is the probability that the alignment was produced by chance. For a single alignment, the e value can be calculated according to Karlin et al., Proc. Natl. Acad. Sci. (1990) 87:2264 and Karlin et al., Proc. Natl. Acad. Sci. (1993) 90. The e value of multiple alignments using the same query sequence can be calculated using an heuristic approach described in Altschul et al., Nat. Genet. (1994) 6:119. Alignment programs such as BLAST program can calculate the e value.

Another factor to consider for determining identity or similarity is the location of the similarity or identity. Strong local alignment can indicate similarity even if the length of alignment is short. Sequence identity scattered throughout the length of the query sequence also can indicate a similarity between the query and profile sequences. The boundaries of the region where the sequences align can be determined according to Doolittle, *supra*; BLAST or FASTA programs; or by determining the area where sequence identity is highest.

In general, in alignment results considered to be of high similarity, the percent of the alignment region length is typically at least about 55% of total length query sequence; more typically, at least about 58%; even more typically; at least about 60% of the total residue length of the query sequence. Usually, percent length of the

alignment region can be as much as about 62%; more usually, as much as about 64%; even more usually, as much as about 66%. Further, for high similarity, the region of alignment, typically, exhibits at least about 75% of sequence identity; more typically, at least about 78%; even more typically; at least about 80% sequence identity. Usually, percent sequence identity can be as much as about 82%; more usually, as much as about 84%; even more usually, as much as about 86%.

The p value is used in conjunction with these methods. The query sequence is considered to have a high similarity with a profile sequence when the p value is less than or equal to 10^{-2} . Confidence in the degree of similarity between the query sequence and the profile sequence increases as the p value become smaller.

In general, where alignment results considered to be of weak similarity, there is no minimum percent length of the alignment region nor minimum length of alignment.

A better showing of weak similarity is considered when the region of alignment is, typically, at least about 15 amino acid residues in length; more typically, at least about 20; even more typically; at least about 25 amino acid residues in length. Usually, length of the alignment region can be as much as about 30 amino acid residues; more usually, as much as about 40; even more usually, as much as about 60 amino acid residues.

Further, for weak similarity, the region of alignment, typically, exhibits at least about 35% of sequence identity; more typically, at least about 40%; even more typically; at least about 45% sequence identity. Usually, percent sequence identity can be as much as about 50%; more usually, as much as about 55%; even more usually, as much as about 60%.

The query sequence is considered to have a low similarity with a profile sequence when the p value is greater than 10^{-2} . Confidence in the degree of similarity between the query sequence and the profile sequence decreases as the p values become larger.

Sequence identity alone can be used to determine similarity of a query sequence to an individual sequence and can indicate the activity of the sequence. Such an alignment, preferably, permits gaps to align sequences. Typically, the query sequence is related to the profile sequence if the sequence identity over the entire query sequence is at least about 15%; more typically, at least about 20%; even more typically, at least about 25%; even more typically, at least about 50%. Sequence identity alone as a

measure of similarity is most useful when the query sequence is usually, at least 80 residues in length; more usually, 90 residues; even more usually, at least 95 amino acid residues in length. More typically, similarity can be concluded based on sequence identity alone when the query sequence is preferably 100 residues in length; more preferably, 120 residues in length; even more preferably, 150 amino acid residues in length.

It is apparent, when studying protein sequence families, that some regions have been better conserved than others during evolution. These regions are generally important for the function of a protein and/or for the maintenance of its three-dimensional structure. By analyzing the constant and variable properties of such groups of similar sequences, it is possible to derive a signature for a protein family or domain, which distinguishes its members from all other unrelated proteins. A pertinent analogy is the use of fingerprints by the police for identification purposes. A fingerprint is generally sufficient to identify a given individual. Similarly, a protein signature can be used to assign a new sequence to a specific family of proteins and thus to formulate hypotheses about its function. The PROSITE database is a compendium of such fingerprints (motifs) and may be used with search software such as Wisconsin GCG Motifs to find motifs or fingerprints in query sequences. PROSITE currently contains signatures specific for about a thousand protein families or domains. Each of these signatures comes with documentation providing background information on the structure and function of these proteins (Hofmann *et al.* (1999) Nucleic Acids Res. **27**:215-219; Bucher and Bairoch ., A generalized profile syntax for biomolecular sequences motifs and its function in automatic sequence interpretation (In) ISMB-94; Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology; Altman *et al.* Eds. (1994), pp 53-61, AAAI Press, Menlo Park).

Translations of the provided nucleic acids can be aligned with amino acid profiles that define either protein families or common motifs. Also, translations of the provided nucleic acids can be aligned to multiple sequence alignments (MSA) comprising the polypeptide sequences of members of protein families or motifs. Similarity or identity with profile sequences or MSAs can be used to determine the activity of the gene

products (e.g., polypeptides) encoded by the provided nucleic acids or corresponding cDNA or genes.

Profiles can be designed manually by (1) creating an MSA, which is an alignment of the amino acid sequence of members that belong to the family and (2) constructing a statistical representation of the alignment. Such methods are described, for example, in Birney et al., *Nucl. Acid Res.* (1996) 24(14): 2730-2739. MSAs of some protein families and motifs are available for downloading to a local server. For example, the PFAM database with MSAs of 547 different families and motifs, and the software (HMMER) to search the PFAM database may be downloaded from <ftp://ftp.genetics.wustl.edu/pub/eddy/pfam-4.4/> to allow secure searches on a local server. Pfam is a database of multiple alignments of protein domains or conserved protein regions, which represent evolutionary conserved structure that has implications for the protein's function (Sonnhammer et al. (1998) *Nucl. Acid Res.* 26:320-322; Bateman et al. (1999) *Nucleic Acids Res.* 27:260-262).

The 3D_al_i databank (Pasarella, S. and Argos, P. (1992) *Prot. Engineering* 5:121-137) was constructed to incorporate new protein structural and sequence data. The databank has proved useful in many research fields such as protein sequence and structure analysis and comparison, protein folding, engineering and design and evolution. The collection enhances present protein structural knowledge by merging information from proteins of similar main-chain fold with homologous primary structures taken from large databases of all known sequences. 3D_al_i databank files may be downloaded to a secure local server from http://www.embl-heidelberg.de/argos/ali/ali_form.html.

The identify and function of the gene that correlates to a nucleic acid described herein can be determined by screening the nucleic acids or their corresponding amino acid sequences against profiles of protein families. Such profiles focus on common structural motifs among proteins of each family. Publicly available profiles are known in the art.

In comparing a novel nucleic acid with known sequences, several alignment tools are available. Examples include PileUp, which creates a multiple sequence alignment, and is described in Feng et al., *J. Mol. Evol.* (1987) 25:351. Another method, GAP,

uses the alignment method of Needleman et al., J. Mol. Biol. (1970) 48:443. GAP is best suited for global alignment of sequences. A third method, BestFit, functions by inserting gaps to maximize the number of matches using the local homology algorithm of Smith *et al.* (1981) Adv. Appl. Math. 2:482.

5

IDENTIFICATION OF SECRETED & MEMBRANE-BOUND POLYPEPTIDES

Secreted and membrane-bound polypeptides of the present invention are of interest. Because both secreted and membrane-bound polypeptides comprise a fragment of contiguous hydrophobic amino acids, hydrophobicity predicting algorithms can be used to identify such polypeptides. A signal sequence is usually encoded by both secreted and membrane-bound polypeptide genes to direct a polypeptide to the surface of the cell. The signal sequence usually comprises a stretch of hydrophobic residues. Such signal sequences can fold into helical structures. Membrane-bound polypeptides typically comprise at least one transmembrane region that possesses a stretch of hydrophobic amino acids that can transverse the membrane. Some transmembrane regions also exhibit a helical structure. Hydrophobic fragments within a polypeptide can be identified by using computer algorithms. Such algorithms include Hopp & Woods, Proc. Natl. Acad. Sci. USA (1981) 78:3824-3828; Kyte & Doolittle, J. Mol. Biol. (1982) 157: 105-132; and RAOAR algorithm, Degli Esposti et al., Eur. J. Biochem. (1990) 190: 207-219.

Another method of identifying secreted and membrane-bound polypeptides is to translate the nucleic acids of the invention in all six frames and determine if at least 8 contiguous hydrophobic amino acids are present. Those translated polypeptides with at least 8; more typically, 10; even more typically, 12 contiguous hydrophobic amino acids are considered to be either a putative secreted or membrane bound polypeptide. Hydrophobic amino acids include alanine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, threonine, tryptophan, tyrosine, and valine.

IDENTIFICATION OF THE FUNCTION OF AN EXPRESSION PRODUCT

The biological function of the encoded gene product of the invention may be determined by empirical or deductive methods. One promising avenue, termed

phylogenomics, exploits the use of evolutionary information to facilitate assignment of gene function. The approach is based on the idea that functional predictions can be greatly improved by focusing on how genes became similar in sequence during evolution instead of focusing on the sequence similarity itself. One of the major efficiencies that has emerged from plant genome research to date is that a large percentage of higher plant genes can be assigned some degree of function by comparing them with the sequences of genes of known function.

Alternatively, "reverse genetics" is used to identify gene function. Large collections of insertion mutants are available for *Arabidopsis*, maize, petunia, and snapdragon. These collections can be screened for an insertional inactivation of any gene by using the polymerase chain reaction (PCR) primed with oligonucleotides based on the sequences of the target gene and the insertional mutagen. The presence of an insertion in the target gene is indicated by the presence of a PCR product. By multiplexing DNA samples, hundreds of thousands of lines can be screened and the corresponding mutant plants can be identified with relatively small effort. Analysis of the phenotype and other properties of the corresponding mutant will provide an insight into the function of the gene.

In one method of the invention, the gene function in a transgenic *Arabidopsis* plant is assessed with anti-sense constructs. A high degree of gene duplication is apparent in *Arabidopsis*, and many of the gene duplications in *Arabidopsis* are very tightly linked. Large numbers of transgenic *Arabidopsis* plants can be generated by infecting flowers with *Agrobacterium tumefaciens* containing an insertional mutagen, a method of gene silencing based on producing double-stranded RNA from bidirectional transcription of genes in transgenic plants can be broadly useful for high-throughput gene inactivation (Clough and Bent (1999) Plant J. **17**; Waterhouse *et al.* (1998) Proc. Natl. Acad. Sci. U.S.A. **95**:13959). This method may use promoters that are expressed in only a few cell types or at a particular developmental stage or in response to an external stimulus. This could significantly obviate problems associated with the lethality of some mutations.

Virus-induced gene silencing may also find use for suppressing gene function. This method exploits the fact that some or all plants have a surveillance system that

can specifically recognize viral nucleic acids and mount a sequence-specific suppression of viral RNA accumulation. By inoculating plants with a recombinant virus containing part of a plant gene, it is possible to rapidly silence the endogenous plant gene.

5 Antisense nucleic acids are designed to specifically bind to RNA, resulting in the formation of RNA-DNA or RNA-RNA hybrids, with an arrest of DNA replication, reverse transcription or messenger RNA translation. Antisense nucleic acids based on a selected nucleic acid sequence can interfere with expression of the corresponding gene. Antisense nucleic acids are typically generated within the cell by expression from
10 antisense constructs that contain the antisense strand as the transcribed strand. Antisense nucleic acids based on the disclosed nucleic acids will bind and/or interfere with the translation of mRNA comprising a sequence complementary to the antisense nucleic acid. The expression products of control cells and cells treated with the
15 corresponding to the nucleic acid upon which the antisense construct is based. The protein is isolated and identified using routine biochemical methods.

As an alternative method for identifying function of the gene corresponding to a nucleic acid disclosed herein, dominant negative mutations are readily generated for corresponding proteins that are active as homomultimers. A mutant polypeptide will
20 interact with wild-type polypeptides (made from the other allele) and form a non-functional multimer. Thus, a mutation is in a substrate-binding domain, a catalytic domain, or a cellular localization domain. Preferably, the mutant polypeptide will be overproduced. Point mutations are made that have such an effect. In addition, fusion of different polypeptides of various lengths to the terminus of a protein can yield
25 dominant negative mutants. General strategies are available for making dominant negative mutants (see for example, Herskowitz (1987) Nature 329:219). Such techniques can be used to create loss of function mutations, which are useful for determining protein function.

Another approach for discovering the function of genes utilizes gene chips and
30 microarrays. DNA sequences representing all the genes in an organism can be placed on miniature solid supports and used as hybridization substrates to quantitate the

differentially expressed polypeptide of the invention, as measured by BLAST using the parameters described above. The variant polypeptides can be naturally or non-naturally glycosylated, i.e., the polypeptide has a glycosylation pattern that differs from the glycosylation pattern found in the corresponding naturally occurring protein.

5 In general, the polypeptides of the subject invention are provided in a non-naturally occurring environment, e.g. are separated from their naturally occurring environment. In certain embodiments, the subject protein is present in a composition that is enriched for the protein as compared to a control. As such, purified polypeptide is provided, where by purified is meant that the protein is present in a composition that
10 is substantially free of non-differentially expressed polypeptides, where by substantially free is meant that less than 90%, usually less than 60% and more usually less than 50% of the composition is made up of non-differentially expressed polypeptides.

Also within the scope of the invention are variants; variants of polypeptides include mutants, fragments, and fusions. Mutants can include amino acid substitutions,
15 additions or deletions. The amino acid substitutions can be conservative amino acid substitutions or substitutions to eliminate non-essential amino acids, such as to alter a glycosylation site, a phosphorylation site or an acetylation site, or to minimize misfolding by substitution or deletion of one or more cysteine residues that are not necessary for function. Conservative amino acid substitutions are those that preserve the general
20 charge, hydrophobicity/hydrophilicity, and/or steric bulk of the amino acid substituted.

Variants also include fragments of the polypeptides disclosed herein, particularly biologically active fragments and/or fragments corresponding to functional domains. Fragments of interest will typically be at least about 10 amino acids (aa) to at least about
25 15 aa in length, usually at least about 50 aa in length, and can be as long as 300 aa in length or longer, but will usually not exceed about 1000 aa in length, where the fragment will have a stretch of amino acids that is identical to a polypeptide encoded by a nucleic acid having a sequence of any SEQ ID NOS:1-999, or a homolog thereof.

The protein variants described herein are encoded by nucleic acids that are within the scope of the invention. The genetic code can be used to select the
30 appropriate codons to construct the corresponding variants.

LIBRARIES AND ARRAYS

In general, a library of biopolymers is a collection of sequence information, which information is provided in either biochemical form (e.g., as a collection of nucleic acid or polypeptide molecules), or in electronic form (e.g., as a collection of genetic sequences stored in a computer-readable form, as in a computer system and/or as part of a computer program). The term biopolymer, as used herein, is intended to refer to polypeptides, nucleic acids, and derivatives thereof, which molecules are characterized by the possession of genetic sequences either corresponding to, or encoded by, the sequences set forth in the provided sequence list (seqlist). The sequence information can be used in a variety of ways, e.g., as a resource for gene discovery, as a representation of sequences expressed in a selected cell type, e.g. cell type markers, etc.

The nucleic acid libraries of the subject invention include sequence information of a plurality of nucleic acid sequences, where at least one of the nucleic acids has a sequence of any of SEQ ID NOS:1-999. By plurality is meant one or more, usually at least 2 and can include up to all of SEQ ID NOS:1-999. The length and number of nucleic acids in the library will vary with the nature of the library, e.g., if the library is an oligonucleotide array, a cDNA array, a computer database of the sequence information, etc.

Where the library is an electronic library, the nucleic acid sequence information can be present in a variety of media. "Media" refers to a manufacture, other than an isolated nucleic acid molecule, that contains the sequence information of the present invention. Such a manufacture provides the sequences or a subset thereof in a form that can be examined by means not directly applicable to the sequence as it exists in a nucleic acid. For example, the nucleotide sequence of the present invention, e.g. the nucleic acid sequences of any of the nucleic acids of SEQ ID NOS:1-999, can be recorded on computer readable media, e.g. any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as a floppy disc, a hard disc storage medium, and a magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. One of skill in the

art can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising a recording of the present sequence information. "Recorded" refers to a process for storing information on computer readable medium, using any such methods as known in the art. Any convenient data storage structure can be chosen, based on the means used to access the stored information. A variety of data processor programs and formats can be used for storage, e.g. word processing text file, database format, etc. In addition to the sequence information, electronic versions of the libraries of the invention can be provided in conjunction or connection with other computer-readable information and/or other types of computer-readable files (e.g., searchable files, executable files, etc, including, but not limited to, for example, search program software, etc.)

By providing the nucleotide sequence in computer readable form, the information can be accessed for a variety of purposes. Computer software to access sequence information is publicly available. For example, the BLAST (Altschul et al., supra.) and BLAZE (Brutlag et al. Comp. Chem. (1993) 17:203) search algorithms on a Sybase system can be used identify open reading frames (ORFs) within the genome that contain homology to ORFs from other organisms.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention. The data storage means can comprise any manufacture comprising a recording of the present sequence information as described above, or a memory access means that can access such a manufacture.

"Search means" refers to one or more programs implemented on the computer-based system, to compare a target sequence or target structural motif with the stored sequence information. Search means are used to identify fragments or regions of the genome that match a particular target sequence or target motif. A variety of known algorithms are publicly known and commercially available, e.g. MacPattern

(EMBL), BLASTN, BLASTX (NCBI) and tBLASTX. A "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids, preferably from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues.

5 A "target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration that is formed upon the folding of the target motif, or on consensus sequences of regulatory or active sites. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme
10 active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, hairpin structures, promoter sequences and other expression elements such as binding sites for transcription factors.

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present
15 invention. One format for an output means ranks fragments of the genome possessing varying degrees of homology to a target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences and identifies the degree of sequence similarity contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or
20 target motif with the data storage means to identify sequence fragments of the genome. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer based systems of the present invention.

As discussed above, the "library" of the invention also encompasses biochemical
25 libraries of the nucleic acids of SEQ ID NOS:1-999, e.g., collections of nucleic acids representing the provided nucleic acids. The biochemical libraries can take a variety of forms, e.g. a solution of cDNAs, a pattern of probe nucleic acids stably bound to a surface of a solid support (microarray) and the like. By array is meant an article of manufacture that has a solid support or substrate with one or more nucleic acid targets
30 on one of its surfaces, where the number of distinct nucleic may be in the hundreds, thousand, or tens of thousands. Each nucleic acid will comprise at 18 nt and often at

least 25 nt, and often at least 100 to 1000 nucleotides, and may represent up to a complete coding sequence or cDNA.. A variety of different array formats have been developed and are known to those of skill in the art. The arrays of the subject invention find use in a variety of applications, including gene expression analysis, drug screening, mutation analysis and the like, as disclosed in the above-listed exemplary patent documents.

In addition to the above nucleic acid libraries, analogous libraries of polypeptides are also provided, where the where the polypeptides of the library will represent at least a portion of the polypeptides encoded by SEQ ID NOS:1-999.

GENETICALLY ALTERED CELLS AND TRANSGENICS

The subject nucleic acids can be used to create genetically modified and transgenic organisms, usually plant cells and plants, which may be monocots or dicots.

The term transgenic, as used herein, is defined as an organism into which an exogenous nucleic acid construct has been introduced, generally the exogenous sequences are stably maintained in the genome of the organism. Of particular interest are transgenic organisms where the genomic sequence of germ line cells has been stably altered by introduction of an exogenous construct.

Typically, the transgenic organism is altered in the genetic expression of the introduced nucleotide sequences as compared to the wild-type, or unaltered organism.

For example, constructs that provide for over-expression of a targeted sequence, sometimes referred to as a "knock-in", provide for increased levels of the gene product.

Alternatively, expression of the targeted sequence can be down-regulated or substantially eliminated by introduction of a "knock-out" construct, which may direct transcription of an anti-sense RNA that blocks expression of the naturally occurring mRNA, by deletion of the genomic copy of the targeted sequence, *etc.*

In one method, large numbers of genes are simultaneously introduced in order to explore the genetic basis of complex traits, for example by making plant artificial chromosome (PLAC) libraries. The centromeres in *Arabidopsis* have been mapped and current genome sequencing efforts will extend through these regions. Because *Arabidopsis* telomeres are very similar to those in yeast one may use a hybrid sequence

of alternating plant and yeast sequences that function in both types of organisms, developing yeast artificial chromosome-PLAC libraries, and then introducing them into a suitable plant host to evaluate the phenotypic consequences. By providing a defined chromosomal environment for cloned genes, the use of PLACs may also enhance the ability to produce transgenic plants with defined levels of gene expression.

It has been found in many organisms that there is significant redundancy in the representation of genes in a genome. That is, a particular gene function is likely by represented by multiple copies of similar coding sequences in the genome. These copies are typically conserved in the amino acid sequence, but may diverge in the sequence of non-translated sequences, and in their codon usage. In order to knock out a particular genetic function in an organism, it may not be sufficient to delete a genomic copy of a single gene. In such cases it may be preferable to achieve a genetic knock-out with an anti-sense construct, particularly where the sequence is aligned with the coding portion of the mRNA.

Methods of transforming plant cells are well-known in the art, and include protoplast transformation, tungsten whiskers (Coffee et al., U.S. Pat. No. 5,302,523, issued Apr. 12, 1994), directly by microorganisms with infectious plasmids, use of transposons (U.S. Patent No. 5,792,294), infectious viruses, the use of liposomes, microinjection by mechanical or laser beam methods, by whole chromosomes or chromosome fragments, electroporation, silicon carbide fibers, and microprojectile bombardment.

For example, one may utilize the biolistic bombardment of meristem tissue, at a very early stage of development, and the selective enhancement of transgenic sectors toward genetic homogeneity, in cell layers that contribute to germline transmission.

Biolistics-mediated production of fertile, transgenic maize is described in Gordon-Kamm et al. (1990), Plant Cell 2:603; Fromm et al. (1990) Bio/Technology 8: 833, for example.

Alternatively, one may use a microorganism, including but not limited to, *Agrobacterium tumefaciens* as a vector for transforming the cells, particularly where the targeted plant is a dicotyledonous species. See, for example, U.S. Patent No. 5,635,381. Leung et al. (1990) Curr. Genet. 17(5):409-11 describe integrative transformation of three fertile

hermaphroditic strains of *Arabidopsis thaliana* using plasmids and cosmids that contain an *E. coli* gene linked to *Aspergillus nidulans* regulatory sequences.

Preferred expression cassettes for cereals may include promoters that are known to express exogenous DNAs in corn cells. For example, the Adh1 promoter has been shown to be strongly expressed in callus tissue, root tips, and developing kernels in corn. Promoters that are used to express genes in corn include, but are not limited to, a plant promoter such as the, CaMV 35S promoter (Odell et al., Nature, 313, 810 (1985)), or others such as CaMV 19S (Lawton et al., Plant Mol. Biol., 9, 31F (1987)), nos (Ebert et al., PNAS USA, 84, 5745 (1987)), Adh (Walker et al., PNAS USA, 84, 6624 (1987)), sucrose synthase (Yang et al., PNAS USA, 87, 4144 (1990)), .alpha.-tubulin, ubiquitin, actin (Wang et al., Mol. Cell. Biol., 12, 3399 (1992)), cab (Sullivan et al., Mol. Gen. Genet, 215, 431 (1989)), PEPCase (Hudspeth et al., Plant Mol. Biol., 12, 579 (1989)), or those associated with the R gene complex (Chandler et al., The Plant Cell, 1, 1175 (1989)). Other promoters useful in the practice of the invention are known to those of skill in the art.

Tissue-specific promoters, including but not limited to, root-cell promoters (Conkling et al., Plant Physiol., 93, 1203 (1990)), and tissue-specific enhancers (Fromm et al., The Plant Cell, 1, 977 (1989)) are also contemplated to be particularly useful, as are inducible promoters such as water-stress-, ABA- and turgor-inducible promoters (Guerrero et al., Plant Molecular Biology, 15, 11-26)), and the like.

Regulating and/or limiting the expression in specific tissues may be functionally accomplished by introducing a constitutively expressed gene (all tissues) in combination with an antisense gene that is expressed only in those tissues where the gene product is not desired. Expression of an antisense transcript of this preselected DNA segment in an rice grain, using, for example, a zein promoter, would prevent accumulation of the gene product in seed. Hence the protein encoded by the preselected DNA would be present in all tissues except the kernel.

Alternatively, one may wish to obtain novel tissue-specific promoter sequences for use in accordance with the present invention. To achieve this, one may first isolate cDNA clones from the tissue concerned and identify those clones which are expressed specifically in that tissue, for example, using Northern blotting or DNA microarrays.

Ideally, one would like to identify a gene that is not present in a high copy number, but which gene product is relatively abundant in specific tissues. The promoter and control elements of corresponding genomic clones may then be localized using the techniques of molecular biology known to those of skill in the art. Alternatively, promoter elements
5 can be identified using enhancer traps based on T-DNA and/or transposon vector systems (see, for example, Campisi *et al.* (1999) Plant J. **17**:699-707; Gu *et al.* (1998) Development **125**:1509-1517).

In some embodiments of the present invention expression of a DNA segment in a transgenic plant will occur only in a certain time period during the development of the
10 plant. Developmental timing is frequently correlated with tissue specific gene expression. For example, in corn expression of zein storage proteins is initiated in the endosperm about 15 days after pollination.

Ultimately, the most desirable DNA segments for introduction into a plant genome may be homologous genes or gene families which encode a desired trait (e.g.,
15 increased disease resistance) and which are introduced under the control of novel promoters or enhancers, etc., or perhaps even homologous or tissue-specific (e.g., root-, grain- or leaf-specific) promoters or control elements.

The genetically modified cells are screened for the presence of the introduced genetic material. The cells may be used in functional studies, drug screening, *etc.*, e.g.
20 to study chemical mode of action, to determine the effect of a candidate agent on pathogen growth, infection of plant cells, *etc.*

The modified cells are useful in the study of genetic function and regulation, for alteration of the cellular metabolism, and for screening compounds that may affect the biological function of the gene or gene product. For example, a series of small deletions
25 and/or substitutions may be made in the host's native gene to determine the role of different domains and motifs in the biological function. Specific constructs of interest include anti-sense, as previously described, which will reduce or abolish expression, expression of dominant negative mutations, and over-expression of genes.

Where a sequence is introduced, the introduced sequence may be either a
30 complete or partial sequence of a gene native to the host, or may be a complete or partial sequence that is exogenous to the host organism, e.g., an *A. thaliana* sequence

inserted into wheat plants. A detectable marker, such as *aldA*, *lac Z*, *etc.* may be introduced into the locus of interest, where upregulation of expression will result in an easily detected change in phenotype.

One may also provide for expression of the gene or variants thereof in cells or tissues where it is not normally expressed, at levels not normally present in such cells or tissues, or at abnormal times of development, during sporulation, *etc.* By providing expression of the protein in cells in which it is not normally produced, one can induce changes in cell behavior.

DNA constructs for homologous recombination will comprise at least a portion of the provided gene or of a gene native to the species of the host organism, wherein the gene has the desired genetic modification(s), and includes regions of homology to the target locus (see Kempin *et al.* (1997) Nature **389**:802-803). DNA constructs for random integration or episomal maintenance need not include regions of homology to mediate recombination. Conveniently, markers for positive and negative selection are included. Methods for generating cells having targeted gene modifications through homologous recombination are known in the art.

Embodiments of the invention provide processes for enhancing or inhibiting synthesis of a protein in a plant by introducing a provided nucleic acids sequence into a plant cell, where the nucleic acid comprises sequences encoding a protein of interest. For example, enhanced resistance to pathogens may be achieved by inserting a nucleic acid encoding an activator in a vector downstream from a promoter sequence capable of driving constitutive high-level expression in a plant cell. When grown into plants, the transgenic plants exhibit increased synthesis of resistance proteins, and increased resistance to pathogens.

Other embodiments of the invention provide processes for enhancing or inhibiting synthesis of a tolerance factor in a plant by introducing a nucleic acid of the invention into a plant cell, where the nucleic acid comprises sequences encoding a tolerance factor. For example, enhanced tolerance to an environmental stress may be achieved by inserting a nucleic acid encoding an activator in a vector downstream from a promoter sequence capable of driving constitutive high-level expression in a plant cell.

When grown into plants, the transgenic plants exhibit increased synthesis of tolerance proteins, and increased tolerance to environmental stress.

Factors which are involved, directly or indirectly in biosynthetic pathways whose products are of commercial, nutritional, or medicinal value include any factor, usually a protein or peptide, which regulates such a biosynthetic pathway (e.g., an activator or repressor); which is an intermediate in such a biosynthetic pathway; or which is a product that increases the nutritional value of a food product; a medicinal product; or any product of commercial value and/or research interest. Plant and other cells may be genetically modified to enhance a trait of interest, by upregulating or down-regulating factors in a biosynthetic pathway.

SCREENING ASSAYS

The polypeptides encoded by the provided nucleic acid sequences, and cells genetically altered to express such sequences, are useful in a variety of screening assays to determine effect of candidate inhibitors, activators, or modifiers of the gene product. One may determine what insecticides, fungicides and the like have an enhancing or synergistic activity with a gene. Alternatively, one may screen for compounds that mimic the activity of the protein. Similarly, the effect of activating agents may be used to screen for compounds that mimic or enhance the activation of proteins. Candidate inhibitors of a particular gene product are screened by detecting decreased from the targeted gene product.

The screening assays may use purified target macromolecules to screen large compound libraries for inhibitory drugs; or the purified target molecule may be used for a rational drug design program, which requires first determining the structure of the macromolecular target or the structure of the macromolecular target in association with its customary substrate or ligand. This information is then used to design compounds which must be synthesized and tested further. Test results are used to refine the molecular models and drug design process in an iterative fashion until a lead compound emerges.

Drug screening may be performed using an *in vitro* model, a genetically altered cell, or purified protein. One can identify ligands or substrates that bind to, modulate or

mimic the action of the target genetic sequence or its product. A wide variety of assays may be used for this purpose, including labeled *in vitro* protein-protein binding assays, electrophoretic mobility shift assays, immunoassays for protein binding, and the like.

The purified protein may also be used for determination of three-dimensional crystal
5 structure, which can be used for modeling intermolecular interactions.

Where the nucleic acid encodes a factor involved in a biosynthetic pathway, as described above, it may be desirable to identify factors, e.g., protein factors, which interact with such factors. One can identify interacting factors, ligands, substrates that bind to, modulate or mimic the action of the target genetic sequence or its product. A
10 wide variety of assays may be used for this purpose, including labeled *in vitro* protein-protein binding assays, electrophoretic mobility shift assays, immunoassays for protein binding, and the like. *In vivo* assays for protein-protein interactions in *E. coli* and yeast cells are also well-established (see Hu *et al.* (2000) Methods 20:80-94; and Bai and Elledge (1997) Methods Enzymol. 283:141-156).

The purified protein may also be used for determination of three-dimensional
15 crystal structure, which can be used for modeling intermolecular interactions. It may also be of interest to identify agents that modulate the interaction of a factor identified as described above with a factor encoded by a nucleic acid of the invention. Drug screening can be performed to identify such agents. For example, a labeled *in vitro*
20 protein-protein binding assay can be used, which is conducted in the presence and absence of an agent being tested.

The term "agent" as used herein describes any molecule, e.g. protein or pharmaceutical, with the capability of altering or mimicking a physiological function. Generally a plurality of assay mixtures are run in parallel with different agent
25 concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, *i.e.* at zero concentration or below the level of detection.

Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight
30 of more than 50 and less than about 2,500 daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen

bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides.

Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and organism extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries.

Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, *etc.* to produce structural analogs.

Where the screening assay is a binding assay, one or more of the molecules may be joined to a label, where the label can directly or indirectly provide a detectable signal.

Various labels include radioisotopes, fluorescers, chemilumescers, enzymes, specific binding molecules, particles, *e.g.* magnetic particles, and the like. Specific binding molecules include pairs, such as biotin and streptavidin, digoxin and antidigoxin *etc.* For the specific binding members, the complementary member would normally be labeled with a molecule that provides for detection, in accordance with known procedures.

A variety of other reagents may be included in the screening assay. These include reagents like salts, neutral proteins, *e.g.* albumin, detergents, *etc.* that are used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Reagents that improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, *etc.* may be used. The mixture of components are added in any order that provides for the requisite binding. Incubations are performed at any suitable temperature, typically between 4 and 40° C. Incubation

periods are selected for optimum activity, but may also be optimized to facilitate rapid high-throughput screening. Typically between 0.1 and 1 hours will be sufficient.

5 The compounds having the desired biological activity may be administered in an acceptable carrier to a host. The active agents may be administered in a variety of ways. Depending upon the manner of introduction, the compounds may be formulated in a variety of ways. The concentration of therapeutically active compound in the formulation may vary from about 0.01-100 wt.%.

10 It must be noted that as used herein and in the appended claims, the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a complex" includes a plurality of such complexes and reference to "the formulation" includes reference to one or more formulations and equivalents thereof known to those skilled in the art, and so forth.

15 Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs. Although any methods, devices and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods, devices and materials are now described.

20 All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing, for example, the methods and methodologies that are described in the publications which might be used in connection with the presently described invention. The publications discussed above and throughout the text are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention.

25

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the subject invention, and are not intended to limit the scope of what is regarded as the invention.

30 Efforts have been made to ensure accuracy with respect to the numbers used (e.g. amounts, temperature, concentrations, etc.) but some experimental errors and deviations should be allowed for. Unless otherwise indicated, parts are parts by weight,

molecular weight is average molecular weight, temperature is in degrees Celsius, and pressure is at or near atmospheric.

EXPERIMENTAL

5 Cloning and Characterization of *Arabidopsis thaliana* Genes.

Following DNA isolation, sequencing was performed using the Dye Primer Sequencing protocol, below. The sequencing reactions were loaded by hand onto a 48 lane ABI 377 and run on a 36 cm gel with the 36E-2400 run module and extraction. Gel analysis was performed with ABI software.

10 The Phred program was used to read the sequence trace from the ABI sequencer, call the bases and produce a sequence read and a quality score for each base call in the sequence., (Ewing *et al.* (1998) Genome Research 8:175-185; Ewing and Green (1998) Genome Research 8:186-194.) PolyPhred may be used to detect single nucleotide polymorphisms in sequences (Kwok *et al.* (1994) Genomics 25:615-
15 622; Nickerson *et al.* (1997) Nucleic Acids Research 25(14):2745-2751.)

MicroWave Plasmid Protocol: Fill Beckman 96 deep-well growth blocks with 1 ml of TB containing 50 µg of ampicillin per ml. Inoculate each well with a colony picked with a toothpick or a 96-pin tool from a glycerol stock plate. Cover the blocks with a plastic lid
20 and tape at two ends to hold lid in place. Incubate overnight (16-24 hours depending on the host stain) at 37° C with shaking at 275 rpm in a New Brunswick platform shaker. Pellet cells by centrifugation for 20 minutes at 3250 rpm in a Beckman GS-R6K, decant TB and freeze pelleted cell in the 96 well block. Thaw blocks on the bench when ready to continue.

25

Prepare the MW-Tween20 solution

For four blocks:

50ml STET/TWEEN20

30 2 tubes RNase (10mg/ml,600ulea)

1 tube lysozyme (25mg)

For 16 blocks:

200ml STET/TWEEN

8 tubes RNase

4 tubes lysozyme

Pipette RNase and Lysozyme into the corner of a beaker. Add Tween 20 solution and swirl to mix completely. Use the Multidrop (or Biohit) to add 25ul of sterile H₂O (from the L size autoclaved bottles) to each well. Resuspend the pellets by vortexing on setting 10 of the platform vortexer. Check pellets after 4 min. and repeat as necessary to resuspend completely. Use the multidrop to add 70 μ l of the freshly prepared MW-Tween 20 solution to each well. Vortex at setting 6 on the platform vortex for 15 seconds. Do not cause frothing.

Incubate the blocks at room temperature for 5 min. Place two blocks at a time in the microwave (1000 Watts) with the tape (placed on the H1 to H12 side of the block) facing away from each other and turn on at full power for 30 seconds. Rotate the blocks so that the tapes face towards each other and turn on at full power again for 30 seconds.

Immediately remove the blocks from the microwave and add 300 μ l of sterile ice cold H₂O with the Multidrop. Seal the blocks with foil tape and place them in an H₂O/ice bath.

Vortex the blocks on 5 for 15 seconds and leave them in the H₂O/Ice bath. Return to step 7 until all the blocks are in the ice water bath. Incubate the blocks for 15 minutes on ice. Spin the blocks for 30 minutes in the Beckman GS-6KR with GH3.8 rotor with Microplus carrier at 3250rpm.

Transfer 100 μ l of the supernatant to Corning/Costar round bottom 96 well trays. Cover with foil and put into fridge if to be sequenced right away. If not to be sequenced in the next day, freeze them at -20° C.

Dye Primer Sequencing: Spin down the DP brew trays and DNA template by pulsing in the Beckman GS-6KR with GH3.8 rotor with Microplus carrier. Big Dye Primer reaction mix trays (one 96 well cycleplate (Robbins) for each nucleotide), 3 microliters of reaction mix per well.

Use twelve channel pipetter (Costar) to add 2 μ l of template to one each G,A,T,C, trays for each template plate. Pulse again to get both the reaction mix and template into the bottom of the cycle plate and put them into the MJ Research DNA Tetrad (PTC-225).

Start program Dye-Primer. Dye-primer is:

96° C, 1 min 1 cycle

96° C, 10 sec.

55° C, 5 sec.

5 70° C, 1 min 15 cycles

96° C, 10 sec.

70° C, 1 min. 15 cycles

4° C soak

When done cycling, using the Robbins Hydra 290 add 100 µl of 100 % ethanol to the

10 A reaction cycle plate and pool the contents of all four cycle plates into the appropriate well.

To perform ethanol precipitation: Use Hydra program 4 to add 100 µl 100% ethanol to each A tray. Use Hydra program 5 to transfer the ethanol and therefore combine the samples from plate to plate. Once the G, A, T, and C trays of each block
15 are mixed, spin for 30 minutes at 3250 in the Beckman. Pour off the ethanol with a firm shake and blot on a paper towel before drying in the speed vac (~10 minutes or until dry). If ready to load add 3 µl dye and denature in the oven at 95° C for ~5 minutes and load 2 µl. If to store, cover with tape and store at -20°C.

20 Common Solutions

Terrific Broth

Per liter:

900 ml H₂O

12 g bacto tryptone

25 24 g bacto-yeast extract

4 ml glycerol

Shake until dissolved and then autoclave. Allow the solution to cool to 60° C or less and then add 100 ml of sterile 0.17M KH₂PO₄, 0.72M K₂HPO₄ (in the hood w/ sterile
30 technique).

0.17M KH₂PO₄, 0.72M K₂HPO₄

Dissolve 2.31g of KH_2PO_4 and 12.54g of K_2HPO_4 in 90 ml of H_2O .

Adjust volume to 100 ml with H_2O and autoclave.

Sequence loading Dye

20 ml deionized formamide

5 3.6 ml dH_2O

400 μl 0.5M EDTA, pH 8.0

0.2 g Blue Dextran

*Light sensitive, cover in foil or store in the dark.

10 STET/TWEEN

10 ml 5M NaCl

5 ml 1M Tris, pH 8.0

1 ml 0.5M EDTA., pH 8.0

25ml Tween20

15 Bring volume to 500 ml with H_2O

The sequencing reactions are run on an ABI 377 sequencer per manufacturer's instructions. The sequencing information obtained each run are analyzed as follows.

Sequencing reads are screened for ribosomal., mitochondrial., chloroplast or human sequence contamination.. In good sequences, vector is marked by x's. These sequences go into biolims regardless of whether or not they pass the criteria for a 'good' sequence. This criteria is ≥ 100 bases with phred score of ≥ 20 and 15 of these bases adjacent to each other.

Sequencing reads that pass the criteria for good sequences are downloaded for assembly into consensus sequences (contigs). The program Phrap (copyrighted by Phil Green at University of Washington, Seattle, WA) utilizes both the Phred sequence information and the quality calls to assemble the sequencing reads. Parameters used with Phrap were determined empirically to minimize assembly of chimeric sequences and maximize differential detection of closely related members of gene families. The following parameters were used with the Phrap program to perform the assembly:

Penalty	-6	Penalty for mismatches(substitutions)
Minmatch	40	Minimum length of matching sequence to use in assembly of

		reads
Trim penalty	0	penalty used for identifying degenerate sequence at beginning and end of read.
Minscore	80	Minimum alignment score

Results from the Phrap analysis yield either contigs consisting of a consensus of two or more overlapping sequence reads, or singlets that are non-overlapping .

The contig and singlets assembly were further analyzed to eliminate low quality sequence utilizing a program to filter sequences based on quality scores generated by the Phred program. The threshold quality for "high quality" base calls is 20. Sequences with less than 50 contiguous high quality bases calls at the beginning of the sequence, and also at the end of the sequence were discarded. Additionally, the maximum allowable percentage of "low quality base calls in the final sequence is 2%, otherwise the sequence is discarded.

The stand-alone BLAST programs and Genbank databases were downloaded from NCBI for use on secure servers at the Paradigm Genetics, Inc. site. The sequences from the assembly were compared to the GenBank NR database downloaded from NCBI using the gapped version (2.0) of BLASTX. BLASTX translates the DNA sequence in all six reading frames and compares it to an amino acid database.

Low complexity sequences are filtered in the query sequence. (Altschul *et al.* (1997) Nucleic Acids Res **25**(17):3389-402).

Genbank sequences found in the BLASTX search with an E Value of less than $1e^{-10}$ are considered to be highly similar, and the Genbank definition lines were used to annotate the query sequences.

When no significantly similar sequences were found as a result of the BLASTX search, the query sequences were compared with the PROSITE database (Bairoch, A. (1992) PROSITE: A dictionary of sites and patterns in proteins. *Nucleic Acids Research* **20**:2013-2018.) to locate functional motifs.

Query sequences were first translated in six reading frames using the Wisconsin GCG pepdata program (Wisconsin Package Version 10.0, Genetics Computer Group (GCG) , Madison, Wisconsin, USA.). The Wisconsin GCG motifs program (Wisconsin Package Version 10.0, Genetics Computer Group (GCG) , Madison, Wisconsin, USA.)

was used to locate motifs in the peptide sequence, with no mismatches allowed. Motif names from the PROSITE results were used to annotate these query sequences.

Table 1

SEQ ID	Reference	Annotation
1	2023001	0 >emb CAB10331.1 (Z97339) pyruvate, orthophosphate dikinase [Arabidopsis thaliana] Length = 960
2	2023002	1E-169 >sp O02654 ENO_LOLPE ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE) (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE) >gi 1911573 bbs 175625 (S80961) enolase [Loligo pealii=squids, nervous system, Peptide, 434 aa] [Loligo pealei] Length = 434
3	2023003	0 >gi 1669387 (U41998) actin 2 [Arabidopsis thaliana] Length = 377
4	2023004	1E-10 >sp P44677 TOLB_HAEIN TOLB PROTEIN PRECURSOR >gi 1073946 pir F64064 colicin tolerance protein (tolB) homolog - Haemophilus influenzae (strain Rd KW20) >gi 1573352 (U32722) colicin tolerance protein (tolB) [Haemophilus influenzae Rd] Length = 427
5	2023005	0 >gi 2062164 (AC001645) jasmonate inducible protein isolog [Arabidopsis thaliana] Length = 470
6	2023006	0 >emb CAA20523.1 (AL031369) Protein phosphatase 2C-like protein [Arabidopsis thaliana] >gi 4559345 gb AAD23006.1 AC006585_1 (AC006585) protein phosphatase 2C [Arabidopsis thaliana] Length = 355
7	2023007	0 >sp P31167 ADT1_ARATH ADP,ATP CARRIER PROTEIN 1 PRECURSOR (ADP/ATP TRANSLOCASE 1) (ADENINE NUCLEOTIDE TRANSLOCATOR 1) (ANT 1) >gi 99658 pir S21313 ADP,ATP carrier protein - Arabidopsis thaliana (fragment) >gi 16175 emb CAA46518 (X65549) adenylate translocator [Arabidopsis thaliana] >gi 445607 prf 1909354A adenylate translocator [Arabidopsis thaliana] Length = 379
8	2023008	0 >sp P29517 TBB9_ARATH TUBULIN BETA-9 CHAIN >gi 320190 pir JQ1593 tubulin beta-9 chain - Arabidopsis thaliana >gi 166910 (M84706) beta-9 tubulin [Arabidopsis thaliana] >gi 5262779 emb CAB45884.1 (AL080282) tubulin beta-9 chain [Arabidopsis thaliana] Length = 444
9	2023009	0 >pir S71288 magnesium chelatase chain - Arabidopsis thaliana >gi 1154627 emb CAA92802 (Z68495) magnesium chelatase subunit [Arabidopsis thaliana] Length = 1381
10	2023010	1E-133 >sp P92966 RS41_ARATH ARGININE/SERINE-RICH SPLICING FACTOR RSP41 >gi 1707370 emb CAA67799 (X99436) splicing factor [Arabidopsis thaliana] Length = 356
11	2023011	0 >dbj BAA11682 (D83025) proline oxidase precursor [Arabidopsis thaliana] Length = 499
12	2023012	0 >sp P17614 ATP2_NICPL ATP SYNTHASE BETA CHAIN, MITOCHONDRIAL PRECURSOR >gi 82133 pir A24355 H+-transporting ATP synthase (EC 3.6.1.34) beta-1 chain, mitochondrial - curled-leaved tobacco >gi 19685 emb CAA26620 (X02868) ATP synthase beta subunit [Nicotiana glauca] Length = 560
13	2023013	0 >gi 2160158 (AC000132) Similar to elongation factor 1-gamma (gb EF1G_XENLA). ESTs gb T20564,gb T45940,gb T04527 come from this gene. [Arabidopsis thaliana] Length = 414
14	2023014	Rgd(2092-2094)
15	2023015	0 >sp P49676 BGAL_BRAOL BETA-GALACTOSIDASE PRECURSOR (LACTASE) >gi 1076460 pir S52393 beta-galactosidase (EC 3.2.1.23) - wild cabbage >gi 669059 emb CAA59162 (X84684) beta-galactosidase [Brassica oleracea] Length = 828
16	2023016	0 >pir S08534 translation elongation factor eEF-1 alpha chain (gene A4)

		- Arabidopsis thaliana >gi 295789 emb CAA34456 (X16432) elongation factor 1-alpha [Arabidopsis thaliana] Length = 449
17	2023017	2E-68 >gi 4091806 (AF052585) CONSTANS-like protein 2 [Malus domestica] Length = 329
18	2023018	0 >sp O24456 GBLP_ARATH GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKE PROTEIN (WD-40 REPEAT AUXIN-DEPENDENT PROTEIN ARCA) >gi 2289095 (U77381) WD-40 repeat protein [Arabidopsis thaliana] Length = 327
19	2023019	1E-140 >sp Q03460 GLSN_MEDSA GLUTAMATE SYNTHASE [NADH] PRECURSOR (NADH-GOGAT) >gi 484529 pir JQ1977 glutamate synthase (NADH) (EC 1.4.1.14) - alfalfa >gi 166412 (L01660) NADH-glutamate synthase [Medicago sativa] Length = 2194
20	2023020	1E-159 >gi 2677828 (U93166) cysteine protease [Prunus armeniaca] Length = 358
21	2023021	3E-74 >sp P31167 ADT1_ARATH ADP,ATP CARRIER PROTEIN 1 PRECURSOR (ADP/ATP TRANSLOCASE 1) (ADENINE NUCLEOTIDE TRANSLOCATOR 1) (ANT 1) >gi 99658 pir S21313 ADP,ATP carrier protein - Arabidopsis thaliana (fragment) >gi 16175 emb CAA46518 (X65549) adenylate translocator [Arabidopsis thaliana] >gi 445607 prf 1909354A adenylate translocator [Arabidopsis thaliana] Length = 379
22	2023022	1E-136 >pir S71265 ferritin - Arabidopsis thaliana >gi 1246401 emb CAA63932 (X94248) ferritin [Arabidopsis thaliana] Length = 255
23	2023023	0 >sp P25856 G3PA_ARATH GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE A, CHLOROPLAST PRECURSOR >gi 2117520 pir JQ1285 glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating) (EC 1.2.1.13) A precursor, chloroplast - Arabidopsis thaliana >gi 166704 (M64117) glyceraldehyde 3-phosphate dehydrogenase [Arabidopsis thaliana] >gi 1402885 emb CAA66816 (X98130) glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating) [Arabidopsis thaliana] Length = 396
24	2023024	Tyr Phospho Site(1382-1388)
25	2023025	0 >gi 2062167 (AC001645) Proline-rich protein APG isolog [Arabidopsis thaliana] Length = 322
26	2023026	0 >gi 3834314 (AC005679) Similar to gene pi010 glycosyltransferase gi 2257490 from S. pombe clone 1750 gb AB004534. ESTs gb T46079 and gb AA394466 come from this gene. [Arabidopsis thaliana] Length = 405
27	2023027	0 >sp P25856 G3PA_ARATH GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE A, CHLOROPLAST PRECURSOR >gi 2117520 pir JQ1285 glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating) (EC 1.2.1.13) A precursor, chloroplast - Arabidopsis thaliana >gi 166704 (M64117) glyceraldehyde 3-phosphate dehydrogenase [Arabidopsis thaliana] >gi 1402885 emb CAA66816 (X98130) glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating) [Arabidopsis thaliana] Length = 396
28	2023028	1E-170 >pir UQMUM ubiquitin precursor - Arabidopsis thaliana >gi 17678 emb CAA31331 (X12853) polyubiquitin (AA 1 - 382) [Arabidopsis thaliana] >gi 987519 (U33014) polyubiquitin [Arabidopsis thaliana] >gi 226499 prf 1515347A poly-ubiquitin [Arabidopsis thaliana] Length = 382
29	2023029	3E-71 >sp P37707 B2_DAUCA B2 PROTEIN >gi 322726 pir S32124 B2 protein - carrot >gi 297889 emb CAA51078 (X72385) B2 protein [Daucus carota] Length = 207
30	2023030	0 >sp P49078 ASNS_ARATH ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE) >gi 507946 (L29083) glutamine-dependent asparagine synthetase [Arabidopsis thaliana] >gi 5541701 emb CAB51206.1 (AL096860) glutamine-dependent asparagine synthetase [Arabidopsis thaliana] Length = 584
31	2023031	2E-25 >gb AAD24193.1 AF134238_1 (AF134238) PL6 protein [Mus musculus] Length = 350

32	2023032	1E-149 >sp P04778 CB22_ARATH CHLOROPHYLL A-B BINDING PROTEIN 2 PRECURSOR (LHCII TYPE I CAB-2) (CAB-140) (LHCP) >gi 16376 emb CAA27543 (X03909) chlorophyll a/b binding protein (LHCP AB 140) [Arabidopsis thaliana] Length = 267
33	2023033	1E-153 >gi 1915974 (U62329) fructokinase [Lycopersicon esculentum] >gi 2102693 (U64818) fructokinase [Lycopersicon esculentum] Length = 328
34	2023034	1E-106 >sp Q64516 GLPK_MOUSE GLYCEROL KINASE (ATP:GLYCEROL 3-PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK) >gi 1480469 (U48403) glycerol kinase [Mus musculus] Length = 524
35	2023035	1E-103 >emb CAA16745.1 (AL021711) heat shock transcription factor-like protein [Arabidopsis thaliana] Length = 401
36	2023036	1E-170 >gi 2286153 (AF007581) cytoplasmic malate dehydrogenase [Zea mays] Length = 332
37	2023037	Tyr_Phospho_Site(1338-1344)
38	2023038	1E-179 >sp P19456 PMA2_ARATH PLASMA MEMBRANE ATPASE 2 (PROTON PUMP) >gi 67973 pir PXMUP2 H+-transporting ATPase (EC 3.6.1.35) type 2, plasma membrane - Arabidopsis thaliana >gi 166629 (J05570) H+-ATPase [Arabidopsis thaliana] >gi 5730129 emb CAB52463.1 (AL109796) H+-transporting ATPase type 2, plasma membrane [Arabidopsis thaliana] Length = 948
39	2023039	Pkc_Phospho_Site(5-7)
40	2023040	Tyr_Phospho_Site(830-837)
41	2023041	8E-98 >gi 4204274 (AC004146) ribulose biphosphate carboxylase, small subunit [Arabidopsis thaliana] Length = 180
42	2023042	1E-175 >emb CAB38206 (AL035601) auxin-responsive GH3-like protein [Arabidopsis thaliana] Length = 603
43	2023043	Pkc_Phospho_Site(19-21)
44	2023044	9E-58 >sp P26599 PTB_HUMAN POLYPYRIMIDINE TRACT-BINDING PROTEIN (PTB) (HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN I) (HNRNP I) (57 KD RNA-BINDING PROTEIN PPTB-1) >gi 35768 emb CAA43973 (X62006) polypirimidine tract binding protein [Homo sapiens] >gi 35774 emb CAA43056 (X60648) polypyrimidine tract-binding protein (pPTB) [Homo sapiens] >gi 4096061 (AC006273) PTB_HUMAN; PTB; HETEROGENEOUS NUCLEA; HNRNP I; 57 KD RNA-BINDING PROTEIN PPTB-1 [Homo sapiens] Length = 531
45	2023045	2E-79 >gi 2642429 (AC002391) poly(A)-binding protein [Arabidopsis thaliana] Length = 662
46	2023046	0 >sp Q38854 CLA1_ARATH PROBABLE 1-DEOXYXYLULOSE-5-PHOSPHATE SYNTHASE PRECURSOR (DXP SYNTHASE) >gi 1399261 (U27099) DEF [Arabidopsis thaliana] Length = 717
47	2023047	Wd_Repeats(1245-1259)
48	2023048	1E-151 >dbj BAA25181 (D88537) delta 9 desaturase [Arabidopsis thaliana] Length = 307
49	2023049	1E-167 >emb CAB43488.1 (AJ012278) ATP-dependent Clp protease subunit ClpP [Arabidopsis thaliana] >gi 5360579 dbj BAA82065.1 (AB022326) nClpP1 [Arabidopsis thaliana] Length = 298
50	2023050	0 >emb CAA67339 (X98807) peroxidase ATP21a [Arabidopsis thaliana] Length = 329
51	2023051	0 >gb AAD39650.1 AC007591_15 (AC007591) Similar to gb Z70524 PDR5-like ABC transporter from Spirodela polyrrhiza and is a member of the PF 00005 ABC transporter family. ESTs gb N97039 and gb T43169 come from this gene. [Arabid
52	2023052	5E-52 >sp P41227 ARDH_HUMAN N-TERMINAL ACETYLTRANSFERASE COMPLEX ARD1 SUBUNIT HOMOLOG >gi 517485 emb CAA54691 (X77588) ARD1 N-acetyl transferase homologue [Homo sapiens] >gi 1302661 (U52112) ARD1 N-acetyl transferase related protein [Homo sapiens] Length = 235
53	2023053	1E-126 >gi 3158476 (AF067185) aquaporin 2 [Samanea saman] Length

		= 287
54	2023054	1E-173 >gi 3212877 (AC004005) Lea-like protein [Arabidopsis thaliana] Length = 325
55	2023055	1E-14 >sp Q28955 PNAD_PIG PROTEIN N-TERMINAL ASPARAGINE AMIDOHYDROLASE (PROTEIN NH2-TERMINAL ASPARAGINE DEAMIDASE) (NTN-AMIDASE) (PNAD) (PROTEIN NH2-TERMINAL ASPARAGINE AMIDOHYDROLASE) (PNAA) >gi 1082956 pir A55768 asparaginyl-peptide amidohydrolase (EC 3.5.1.-) - pig >gi 595950 (U17062) protein N-terminal asparagine amidohydrolase [Sus scrofa] Length = 310
56	2023056	1E-172 >sp P53799 FDFT_ARATH FARNESYL-DIPHOSPHATE FARNESYLTRANSFERASE (SQUALENE SYNTHETASE) (SQS) (SS) (FPP:FPP FARNESYLTRANSFERASE) >gi 1076324 pir S54251 farnesyl-diphosphate farnesyltransferase (EC 2.5.1.21) - Arabidopsis thaliana >gi 798820 emb CAA60385 (X86692) farnesyl-diphosphate farnesyltransferase [Arabidopsis thaliana] >gi 806325 dbj BAA06103 (D29017) squalene synthase [Arabidopsis thaliana] >gi 2232212 (AF004560) squalene synthase 1 [Arabidopsis thaliana] >gi 3096933 emb CAA18843.1 (AL023094) farnesyl-diphosphate farnesyltransferase [Arabidopsis thaliana] >gi 4098519 (U79159) squalene synthase [Arabidopsis thaliana] Length = 410
57	2023057	1E-141 >gi 3413700 (AC004747) YME1 protein [Arabidopsis thaliana] Length = 627
58	2023058	Tyr_Phospho_Site(1667-1673)
59	2023059	1E-144 >sp Q08682 RSP4_ARATH 40S RIBOSOMAL PROTEIN SA (P40) (LAMININ RECEPTOR HOMOLOG) >gi 322536 pir S30570 laminin receptor homolog - Arabidopsis thaliana >gi 16380 emb CAA48794 (X69056) laminin receptor homologue [Arabidopsis thaliana] Length = 298
60	2023060	2E-43 >gi 2735550 (U96638) unc-50 related protein; URP [Rattus norvegicus] Length = 259
61	2023061	Tyr_Phospho_Site(65-73)
62	2023062	2E-30 >emb CAB03470.1 (Z81137) Similarity to Yeast YIP1 protein (SW:P53039); cDNA EST EMBL:T01608 comes from this gene; cDNA EST EMBL:C07393 comes from this gene; cDNA EST EMBL:C07550 comes from this gene; cDNA EST EMBL:C08746 comes from this gene... Length = 282
63	2023063	1E-151 >gi 1773330 (U80071) glycolate oxidase [Mesembryanthemum crystallinum] Length = 370
64	2023064	7E-44 >ref NP_006339.1 PGTC90 Golgi transport complex protein (90 kDa) >gi 3808235 (AF058718) 13 S Golgi transport complex 90kD subunit brain-specific isoform [Homo sapiens] Length = 839
65	2023065	1E-168 >emb CAB44681.1 (AL078620) mitochondrial carrier-like protein [Arabidopsis thaliana] Length = 330
66	2023066	4E-12 >gi 1764100 (U81805) GDP-D-mannose-4,6-dehydratase [Arabidopsis thaliana] Length = 373
67	2023067	2E-22 >gb AAD48936.1 AF160760_4 (AF160760) contains similarity to Pfam family PF0040 - WD domain, G-beta repeat; score=10.8, E=3.2, N=2 [Arabidopsis thaliana] Length = 892
68	2023068	1E-123 >sp P30302 WC2C_ARATH PLASMA MEMBRANE INTRINSIC PROTEIN 2C (WATER-STRESS INDUCED TONOPLAST INTRINSIC PROTEIN) (WSI-TIP) >gi 217869 dbj BAA02520 (D13254) transmembrane channel protein [Arabidopsis thaliana] >gi 4371283 gb AAD18141 (AC006260) plasma membrane intrinsic protein 2C [Arabidopsis thaliana] >gi 384324 prf 1905411A transmembrane channel [Arabidopsis thaliana] Length = 285
69	2023069	6E-12 >dbj BAA74463 (AB022605) mRNA (guanine-7-)methyltransferase [Homo sapiens] Length = 504
70	2023070	1E-153 >gi 2062161 (AC001645) jasmonate inducible protein isolog [Arabidopsis thaliana] Length = 298
71	2023071	1E-157 >sp P43286 WC2A_ARATH PLASMA MEMBRANE INTRINSIC

		PROTEIN 2A >gi 629542 pir S44084 plasma membrane intrinsic protein 2a - Arabidopsis thaliana >gi 472877 emb CAA53477 (X75883) plasma membrane intrinsic protein 2a [Arabidopsis thaliana] Length = 287
72	2023072	9E-98 >gi 2252840 (AF013293) contains regions of similarity to Haemophilus influenzae permease (SP:P38767) [Arabidopsis thaliana] >gi 6049882 gb AAF02797.1 AF195115_17 (AF195115) contains regions of similarity to Haemophilus influenzae permease (SP:P38767) [Arabidopsis thaliana] Length = 746
73	2023073	9E-97 >gb AAF00673.1 AC008153_25 (AC008153) 2-cys peroxiredoxin BAS1 precursor (thiol-specific antioxidant protein) [Arabidopsis thaliana] >gi 6041816 gb AAF02131.1 AC009918_3 (AC009918) 2-cys peroxiredoxin [Arabidopsis thaliana] Length = 301
74	2023074	1E-168 >emb CAA06460 (AJ005261) cytidine deaminase [Arabidopsis thaliana] >gi 3093276 emb CAA06671.1 (AJ005687) cytidine deaminase [Arabidopsis thaliana] >gi 4191787 (AC005917) cytidine deaminase [Arabidopsis thaliana] >gi 6090835 gb AAF03358.1 AF134487_1 (AF134487) cytidine deaminase 1 [Arabidopsis thaliana] Length = 301
75	2023075	0 >emb CAA66863 (X98190) peroxidase ATP2a [Arabidopsis thaliana] >gi 4371288 gb AAD18146 (AC006260) peroxidase ATP2a [Arabidopsis thaliana] Length = 327
76	2023076	1E-159 >sp Q08733 WC1C_ARATH PLASMA MEMBRANE INTRINSIC PROTEIN 1C (TRANSMEMBRANE PROTEIN B) (TMP-B) >gi 396218 emb CAA49155 (X69294) transmembrane protein TMP-B [Arabidopsis thaliana] Length = 286
77	2023077	Rgd(840-842)
78	2023078	1E-157 >emb CAB10405.1 (Z97340) beta-1, 3-glucanase class I precursor [Arabidopsis thaliana] Length = 306
79	2023079	1E-110 >gi 3341679 (AC003672) dynamin-like protein phragmoplastin 12 [Arabidopsis thaliana] Length = 613
80	2023080	1E-79 >gb AAA02747.1 (L13655) membrane protein [Saccharum hybrid cultivar H65-7052] Length = 325
81	2023081	1E-155 >pir S33443 chlorophyll a/b-binding protein CP29 - Arabidopsis thaliana >gi 298036 emb CAA50712 (X71878) CP29 [Arabidopsis thaliana] Length = 290
82	2023082	0 >emb CAB56580.1 (AJ011628) squamosa promoter binding protein-like 1 [Arabidopsis thaliana] Length = 881
83	2023083	Tyr Phospho Site(305-312)
84	2023084	6E-22 >gb AAD46141.1 AF081022_1 (AF081022) hypoxia-induced protein kinase L31 [Lycopersicon esculentum] Length = 78
85	2023085	1E-154 >gi 2281109 (AC002333) endochitinase isolog [Arabidopsis thaliana] Length = 281
86	2023086	1E-79 >gi 3415117 (AF081203) villin 3 [Arabidopsis thaliana] Length = 966
87	2023087	1E-103 >ref NP_005435.1 PRCD1+ protein involved in sexual development >gi 1620898 dbj BAA13508 (D87957) protein involved in sexual development [Homo sapiens] Length = 299
88	2023088	1E-106 >sp Q05047 CP72_CATRO CYTOCHROME P450 72A1 (CYPLXXII) (PROBABLE GERANIOL-10-HYDROXYLASE) (GE10H) >gi 167484 (L10081) Cytochrome P-450 protein [Catharanthus roseus] >gi 445604 prf 1909351A cytochrome P450 [Catharanthus roseus] Length = 524
89	2023089	5E-41 >ref NP_000511.1 PHEXA hexosaminidase A (alpha polypeptide) >gi 123079 sp P06865 HEXA_HUMAN BETA-HEXOSAMINIDASE ALPHA CHAIN PRECURSOR (N-ACETYL-BETA-GLUCOSAMINIDASE) (BETA-N-ACETYLHEXOSAMINIDASE) >gi 67503 pir AOHUBA beta-N-acetylhexosaminidase (EC 3.2.1.52) alpha chain precursor - human >gi 179458 (M16424) beta-hexosaminidase alpha chain [Homo sapiens] >gi 4261632 gb AAD13932 1680052_1 (S62076) lysosomal enzyme beta-N-

		acetylhexosaminidase A [Homo sapiens] Length = 529
90	2023090	0 >emb CAB36796.1 (AL035525) pectinesterase-like protein [Arabidopsis thaliana] Length = 477
91	2023091	1E-139 >emb CAB10240.1 (Z97336) disease resistance RPS2 like protein [Arabidopsis thaliana] Length = 719
92	2023092	1E-170 >pir S49332 seed tetraubiquitin - common sunflower >gi 303901 dbj BAA03764 (D16248) ubiquitin [Glycine max] >gi 456714 dbj BAA05670 (D28123) Ubiquitin [Glycine max] >gi 556688 emb CAA84440 (Z34988) seed tetraubiquitin [Helianthus annuus] >gi 994785 dbj BAA05085 (D26092) Ubiquitin [Glycine max] >gi 4263514 gb AAD15340 (AC004044) polyubiquitin [Arabidopsis thaliana] >gi 1096513 prf 2111434A tetraubiquitin [Helianthus annuus] Length = 305
93	2023093	1E-146 >gi 2088652 (AF002109) 26S proteasome regulatory subunit S12 isolog [Arabidopsis thaliana] >gi 2351376 (U54561) translation initiation factor eIF2 p47 subunit homolog [Arabidopsis thaliana] Length = 293
94	2023094	0 >pir B45511 chitinase (EC 3.2.1.14) precursor, basic - Arabidopsis thaliana >gi 166666 (M38240) basic chitinase [Arabidopsis thaliana] >gi 5689104 dbj BAA82811.1 (AB023449) basic endochitinase [Arabidopsis thaliana] >gi 5689106 dbj BAA82812.1 (AB023450) basic endochitinase [Arabidopsis thaliana] >gi 5689108 dbj BAA82813.1 (AB023451) basic endochitinase [Arabidopsis thaliana] >gi 5689112 dbj BAA82815.1 (AB023453) basic endochitinase [Arabidopsis thaliana] >gi 5689114 dbj BAA82816.1 (AB023454) basic endochitinase [Arabidopsis thaliana] >gi 5689120 dbj BAA82819.1 (AB023457) basic endochitinase [Arabidopsis thaliana] >gi 5689122 dbj BAA82820.1 (AB023458) basic endochitinase [Arabidopsis thaliana] >gi 5689124 dbj BAA82821.1 (AB023459) basic endochitinase [Arabidopsis thaliana] >gi 5689126 dbj BAA82822.1 (AB023460) basic endochitinase [Arabidopsis thaliana] >gi 5689128 dbj BAA82823.1 (AB023461) basic endochitinase [Arabidopsis thaliana] >gi 5689132 dbj BAA82825.1 (AB023463) basic endochitinase [Arabidopsis thaliana] Length = 335
95	2023095	Tyr Phospho Site(1027-1033)
96	2023096	1E-152 >pir S23546 chlorophyll a/b-binding protein type I precursor Lhb1B2 - Arabidopsis thaliana >gi 16364 emb CAA45790 (X64460) photosystem II type I chlorophyll a /b binding protein [Arabidopsis thaliana] >gi 3128230 (AC004077) photosystem II type I chlorophyll a/b binding protein [Arabidopsis thaliana] >gi 3337371 (AC004481) photosystem II type I chlorophyll a/b binding protein [Arabidopsis thaliana] Length = 265
97	2023097	Tyr Phospho Site(98-104)
98	2023098	1E-133 >emb CAB38813.1 (AL035679) ubiquitin-dependent proteolytic protein [Arabidopsis thaliana] Length = 315
99	2023099	5E-45 >gb AAD26911.1 AC006429_1 (AC006429) auxin down-regulated protein [Arabidopsis thaliana] Length = 291
100	2023100	1E-169 >sp P46523 CLPA_BRANA ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA PRECURSOR >gi 480969 pir S37557 clpA protein - rape (fragment) >gi 406311 emb CAA53077 (X75328) clpA [Brassica napus] Length = 874
101	2023101	1E-100 >gb AAD28780.1 AF134133_1 (AF134133) Lil3 protein [Arabidopsis thaliana] Length = 262
102	2023102	4E-42 >gi 3329368 (AF031244) nodulin-like protein [Arabidopsis thaliana] Length = 559
103	2023103	Tyr Phospho Site(206-212)
104	2023104	Tyr Phospho Site(740-748)
105	2023105	1E-130 >pir S20866 L-ascorbate peroxidase (EC 1.11.1.11) precursor - Arabidopsis thaliana (fragment) Length = 263
106	2023106	2E-15 >gi 4093153 (AF088280) phytochrome-associated protein 3

		[Arabidopsis thaliana] Length = 524
107	2023107	Zinc Protease(1292-1301)
108	2023108	1E-148 >dbj BAA32735 (AB011545) GF14 mu [Arabidopsis thaliana] >gi 4559343 gb AAD23005.1 AC007087_24 (AC007087) DNA regulatory protein GF14 mu [Arabidopsis thaliana] >gi 5802796 gb AAD51784.1 AF145301_1 (AF145301) 14-3-3 protein GF14 mu [Arabidopsis thaliana] Length = 263
109	2023109	Zinc Finger C3hc4(138-147)
110	2023110	2E-49 >dbj BAA16833 (D90901) spore germination protein c2 [Synechocystis sp.] Length = 238
111	2023111	2E-44 >emb CAA21916.1 (AL033389) yeast cell division cycle CDC50 homolog [Schizosaccharomyces pombe] Length = 396
112	2023112	Zinc Finger C2h2(879-903)
113	2023113	3E-66 >gb AAD39835.1 AF057024_1 (AF057024) Ran-binding protein siRanBP [Arabidopsis thaliana] Length = 234
114	2023114	1E-173 >gb AAD38248.1 AC006193_4 (AC006193) membrane related protein [Arabidopsis thaliana] Length = 385
115	2023115	Tyr Phospho Site(1708-1714)
116	2023116	5E-63 >emb CAA69300 (Y08061) endomembrane-associated protein [Arabidopsis thaliana] >gi 2982443 emb CAA18251.1 (AL022224) endomembrane-associated protein [Arabidopsis thaliana] Length = 225
117	2023117	2E-46 >gi 451193 (L28008) wali7 [Triticum aestivum] >gi 1090845 prf 2019486B wali7 gene [Triticum aestivum] Length = 270
118	2023118	1E-102 >pir S58499 IAA13 protein - Arabidopsis thaliana >gi 972929 (U18415) IAA13 [Arabidopsis thaliana] >gi 2459414 (AC002332) auxin inducible protein, IAA13 [Arabidopsis thaliana] Length = 246
119	2023119	Tyr Phospho Site(14-21)
120	2023120	1E-147 >sp P27521 CB24_ARATH CHLOROPHYLL A-B BINDING PROTEIN 4 PRECURSOR (LHCI TYPE III CAB-4) (LHCP) >gi 166646 (M63931) light- harvesting chlorophyll a/b binding protein [Arabidopsis thaliana] Length = 251
121	2023121	Tyr Phospho Site(414-421)
122	2023122	3E-59 >emb CAB10557.1 (Z97344) trehalose-6-phosphate synthase like protein [Arabidopsis thaliana] Length = 865
123	2023123	Tyr Phospho Site(110-117)
124	2023124	1E-109 >dbj BAA33810.1 (AB018441) phi-1 [Nicotiana tabacum] Length = 313
125	2023125	1E-120 >emb CAB56038.1 (AJ011049) tyrosine decarboxylase [Arabidopsis thaliana] Length = 489
126	2023126	Tyr Phospho Site(640-647)
127	2023127	3E-44 >ref NP_005818.1 PUGTREL1 UDP-galactose transporter related >gi 2136346 pir JC5024 UDP-galactose transporter related isozyme 1 - human >gi 1669560 dbj BAA13525.1 (D87989) UGTrel1 [Homo sapiens] Length = 322
128	2023128	1E-115 >sp P42055 POR4_SOLTU 34 KD OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN (VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN) (VDAC) (POM 34) >gi 629720 pir S46936 34K porin - potato >gi 1076682 pir A55364 porin (clone pPOM-34) - potato mitochondrion >gi 516166 emb CAA56599 (X80386) 34 kDa porin [Solanum tuberosum] Length = 276
129	2023129	Tyr Phospho Site(25-32)
130	2023130	1E-76 >sp Q42656 AGAL_COFA ALPHA-GALACTOSIDASE PRECURSOR (MELIBIASE) (ALPHA-D-GALACTOSIDE GALACTOHYDROLASE) >gi 504489 (L27992) alpha-galactosidase [Coffea arabica] Length = 378
131	2023131	2E-20 >gb AAF01440.1 AF187961_1 (AF187961) ubiquitin carboxyl-terminal hydrolase [Schizosaccharomyces pombe] Length = 1129
132	2023132	1E-141 >emb CAA17567 (AL021961) caffeoyl-CoA O-methyltransferase - like protein [Arabidopsis thaliana] Length = 259

133	2023133	1E-97 >emb CAA64565 (X95269) LRR protein [<i>Lycopersicon esculentum</i>] Length = 221
134	2023134	3E-53 >dbj BAA24576 (AB000778) phospholipase D [<i>Rattus norvegicus</i>] Length = 1074
135	2023135	9E-48 >sp P27061 PPA1_LYCES ACID PHOSPHATASE PRECURSOR 1 >gi 170370 (M83211) acid phosphatase type 1 [<i>Lycopersicon esculentum</i>] >gi 170372 (M67474) acid phosphatase type 5 [<i>Lycopersicon esculentum</i>] >gi 445121 prf 1908427A acid phosphatase 1 [<i>Lycopersicon esculentum</i>] Length = 255
136	2023136	1E-138 >gi 3421072 (AF043519) 20S proteasome subunit PAA2 [<i>Arabidopsis thaliana</i>] >gi 4006819 gb AAC95161.1 (AC005970) 20S proteasome subunit PAA2 [<i>Arabidopsis thaliana</i>] Length = 246
137	2023137	2E-75 >gb AAD14602 (AF092910) stage specific peptide 24 [<i>Trypanosoma cruzi</i>] Length = 287
138	2023138	1E-158 >pir S59519 tryptophan synthase (EC 4.2.1.20) alpha chain - <i>Arabidopsis thaliana</i> >gi 619753 (U18993) tryptophan synthase alpha chain [<i>Arabidopsis thaliana</i>] >gi 1585768 prf 2201482A Trp synthase:SUBUNIT=alpha [<i>Arabidopsis thaliana</i>] Length = 312
139	2023139	Tyr Phospho Site(892-900)
140	2023140	3E-53 >emb CAB43522.1 (AJ238804) non-specific lipid transfer protein [<i>Arabidopsis thaliana</i>] Length = 118
141	2023141	1E-165 >pir S71226 xyloglucan endotransglycosylase-related protein XTR-7 - <i>Arabidopsis thaliana</i> >gi 1244760 (U43489) xyloglucan endotransglycosylase-related protein [<i>Arabidopsis thaliana</i>] Length = 289
142	2023142	1E-146 >gb AAD55272.1 AC008263_3 (AC008263) Identical to gb AF078080 isochorismate synthase from <i>Arabidopsis thaliana</i> . ESTs gb R90272, gb A100274 and gb T42189 come from this gene. Length = 503
143	2023143	1E-158 >sp P43285 WC1A_ARATH PLASMA MEMBRANE INTRINSIC PROTEIN 1A >gi 629540 pir S44082 plasma membrane intrinsic protein 1a - <i>Arabidopsis thaliana</i> >gi 472873 emb CAA53475 (X75881) plasma membrane intrinsic protein 1a [<i>Arabidopsis thaliana</i>] Length = 286
144	2023144	1E-173 >sp Q38882 PLD_ARATH PHOSPHOLIPASE D PRECURSOR (PLD) (CHOLINE PHOSPHATASE) (PHOSPHATIDYLCHOLINE-HYDROLYZING PHOSPHOLIPASE D) >gi 1297302 (U36381) phospholipase D [<i>Arabidopsis thaliana</i>] Length = 809
145	2023145	3E-97 >sp Q03943 IM30_PEA CHLOROPLAST MEMBRANE-ASSOCIATED 30 KD PROTEIN PRECURSOR (M30) >gi 1076532 pir S47966 probable lipid transfer protein M30 precursor - garden pea >gi 169107 (M73744) IM30 [<i>Pisum sativum</i>] Length = 323
146	2023146	1E-167 >sp P55737 HS82_ARATH HEAT SHOCK PROTEIN 81-2 (HSP81-2) >gi 445127 prf 1908431B heat shock protein HSP81-2 [<i>Arabidopsis thaliana</i>] Length = 699
147	2023147	Pkc Phospho Site(56-58)
148	2023148	3E-26 >dbj BAA75919.1 (AB009340) tartrate-resistant acid phosphatase [<i>Oryctolagus cuniculus</i>] Length = 325
149	2023149	1E-159 >emb CAA17774.1 (AL022023) plasma membrane intrinsic protein (SIMIP) [<i>Arabidopsis thaliana</i>] Length = 280
150	2023150	1E-155 >gi 2443883 (AC002294) Similar to RPS-2 disease resistance protein [<i>Arabidopsis thaliana</i>] Length = 967
151	2023151	1E-99 >gb AAD29800.1 AC006264_8 (AC006264) signal sequence receptor, alpha subunit (SSR-alpha) [<i>Arabidopsis thaliana</i>] Length = 257
152	2023152	Tyr Phospho Site(642-650)
153	2023153	2E-64 >gb AAC78271.1 AAC78271 (AC002330) glutamate-/aspartate-binding peptide [<i>Arabidopsis thaliana</i>] Length = 248
154	2023154	1E-172 >gi 4218963 (AF093672) xyloglucan endotransglycosylase

		[Arabidopsis thaliana] >gi 4539300 emb CAB39603.1 (AL049480) xyloglucan endo-1, 4-beta-D-glucanase [Arabidopsis thaliana] Length = 287
155	2023155	Zinc Finger C2h2(917-941)
156	2023156	1E-108 >emb CAA65416 (X96598) CaLB protein [Arabidopsis thaliana] Length = 493
157	2023157	5E-26 >emb CAA64425 (X94976) cell wall-plasma membrane linker protein [Brassica napus] Length = 376
158	2023158	1E-159 >gb AAD25750.1 AC007060_8 (AC007060) Strong similarity to F19I3.7 gi 3033380 coatmer epsilon subunit from Arabidopsis thaliana BAC gb AC004238. ESTs gb Z17908, gb AA728673, gb N96555, gb H76335, gb AA712463, gb W43247, gb T45611, g... Length = 292
159	2023159	Tyr Phospho Site(958-964)
160	2023160	5E-14 >emb CAA18475.1 (AL022347) serine /threonine kinase-like protein, receptor kinase [Arabidopsis thaliana] Length = 656
161	2023161	3E-33 >sp P26568 H11_ARATH HISTONE H1.1 >gi 1070594 pir HSMU11 histone H1.1 - Arabidopsis thaliana >gi 16317 emb CAA44314 (X62458) Histone H1 [Arabidopsis thaliana] Length = 274
162	2023162	2E-97 >emb CAA07573.1 (AJ007586) src2-like protein [Arabidopsis thaliana] Length = 324
163	2023163	Tyr Phospho Site(246-254)
164	2023164	1E-171 >sp Q42547 CAT3_ARATH CATALASE 3 >gi 2347178 (U43147) catalase 3 [Arabidopsis thaliana] >gi 2511726 (AF021937) catalase 3 [Arabidopsis thaliana] Length = 492
165	2023165	Tyr Phospho Site(75-83)
166	2023166	1E-151 >emb CAA66966 (X98322) peroxidase [Arabidopsis thaliana] >gi 1429219 emb CAA67312 (X98776) peroxidase ATP13a [Arabidopsis thaliana] Length = 313
167	2023167	7E-38 >emb CAB41106.1 (AL049656) myb-like protein [Arabidopsis thaliana] Length = 261
168	2023168	8E-74 >gi 4008006 (AF084034) receptor-like protein kinase [Arabidopsis thaliana] Length = 645
169	2023169	1E-137 >pir JQ1678 transcription factor tga1 - Arabidopsis thaliana >gi 16550 emb CAA48189 (X68053) transcription factor [Arabidopsis thaliana] Length = 367
170	2023170	8E-57 >gi 3184559 (AF052290) c-type cytochrome biogenesis protein [Synechococcus PCC7002] Length = 246
171	2023171	1E-103) >gb AAD32768.1 AC007661_5 (AC007661) alpha-carboxyltransferase [Arabidopsis thaliana] Length = 796
172	2023172	1E-117 >gb AAD32822.1 AC007659_4 (AC007659) phosphatidate cytidyltransferase [Arabidopsis thaliana] Length = 430
173	2023173	1E-129 >dbj BAA32210 (AB015138) Vacuolar proton pyrophosphatase [Arabidopsis thaliana] Length = 770
174	2023174	2E-76 >gi 3157927 (AC002131) Contains similarity to GDP-dissociation inhibitor gb L07918 from Mus musculus. [Arabidopsis thaliana] Length = 223
175	2023175	2E-89 >pir S68589 serine/threonine-specific kinase (EC 2.7.1.-) precursor - Arabidopsis thaliana >gi 1405837 emb CAA62824 (X91630) receptor-like kinase [Arabidopsis thaliana] >gi 2150023 (AF001168) receptor-like kinase LECRK1 [Arabidopsis thaliana] Length = 661
176	2023176	7E-86 >gi 3769673 (AF095285) Tic20 [Pisum sativum] Length = 253
177	2023177	2E-17 >sp P46689 GAS1_ARATH GIBBERELLIN-REGULATED PROTEIN 1 PRECURSOR >gi 2129588 pir S71441 GAST1 protein homolog (clone GAS1) - Arabidopsis thaliana >gi 887939 (U11766) GAST1 protein homolog [Arabidopsis thaliana] Length = 98
178	2023178	1E-166 >sp O48661 SPEE_ARATH SPERMIDINE SYNTHASE (PUTRESCINE AMINOPROPYLTRANSFERASE) (SPDSY)

		>gi 2821961 dbj BAA24536 (AB006693) spermidine synthase [Arabidopsis thaliana] Length = 293
179	2023179	Ww_Domain_1(1284-1310)
180	2023180	1E-104 >pir S27762 Sip1 protein - barley >gi 167100 (M77475) seed imbibition protein [Hordeum vulgare] Length = 757
181	2023181	1E-155 >sp P48641 GSHR_ARATH GLUTATHIONE REDUCTASE, CYTOSOLIC (GR) (GRASE) (OBP29) >gi 1022797 (U37697) glutathione reductase [Arabidopsis thaliana] Length = 499
182	2023182	Tyr_Phospho_Site(599-607)
183	2023183	1E-133 >gi 3688799 (AF057137) gamma tonoplast intrinsic protein 2 [Arabidopsis thaliana] Length = 253
184	2023184	1E-110 >gi 3075392 (AC004484) steroid dehydrogenase [Arabidopsis thaliana] Length = 390
185	2023185	Tyr_Phospho_Site(48-56)
186	2023186	6E-38 >emb CAA16875.1 (AL021749) receptor protein kinase like protein [Arabidopsis thaliana] Length = 649
187	2023187	Tyr_Phospho_Site(1737-1743)
188	2023188	1E-128 >sp P48349 143L_ARATH 14-3-3-LIKE PROTEIN GF14 LAMBDA (14-3-3-LIKE PROTEIN AFT1) >gi 1084332 pir S53727 14-3-3-like protein (ATF1) - Arabidopsis thaliana >gi 953221 (U02565) 14-3-3-like protein 1 [Arabidopsis thaliana] >gi 1549404 (U68545) GF14 lambda [Arabidopsis thaliana] >gi 5802790 gb AAD51781.1 AF145298_1 (AF145298) 14-3-3 protein GF14 lambda [Arabidopsis thaliana] Length = 248
189	2023189	1E-135 >emb CAB39932.1 (AL049500) phosphoribosylanthranilate transferase [Arabidopsis thaliana] Length = 857
190	2023190	Serpin(1794-1804)
191	2023191	3E-77 >gi 3319340 (AF077407) contains similarity to E. coli cation transport protein ChaC (GB:D90756) [Arabidopsis thaliana] Length = 197
192	2023192	1E-47 >emb CAA23033.1 (AL035394) major latex protein [Arabidopsis thaliana] Length = 151
193	2023193	7E-76 >gb AAB17191.1 (U73103) laccase [Liriodendron tulipifera] Length = 570
194	2023194	Tyr_Phospho_Site(712-718)
195	2023195	1E-161 >sp Q06611 WC1B_ARATH PLASMA MEMBRANE INTRINSIC PROTEIN 1B (TRANSMEMBRANE PROTEIN A) (TMP-A) >gi 296085 emb CAA48356 (X68293) transmembrane protein [Arabidopsis thaliana] >gi 3386599 (AC004665) plasma membrane intrinsic protein 1B [Arabidopsis thaliana] Length = 286
196	2023196	1E-16 >sp P44445 RLUD_HAEIN RIBOSOMAL LARGE SUBUNIT PSEUDOURIDINE SYNTHASE D (PSEUDOURIDYLATE SYNTHASE) (URACIL HYDROLYASE) >gi 1074296 pir F64144 hypothetical protein HI0176 - Haemophilus influenzae (strain Rd KW20) >gi 1573131 (U32702) sfhB protein (sfhB) [Haemophilus influenzae Rd] Length = 324
197	2023197	2E-22 >gb AAD48964.1 AF147263_6 (AF147263) contains similarity to Medicago truncatula N7 protein (GB:Y17613) [Arabidopsis thaliana] Length = 246
198	2023198	Tyr_Phospho_Site(1422-1428)
199	2023199	Tyr_Phospho_Site(1517-1524)
200	2023200	1E-109 >gi 2642432 (AC002391) elicitor response element binding protein (WRKY3) [Arabidopsis thaliana] Length = 317
201	2023201	Tyr_Phospho_Site(271-279)
202	2023202	1E-176 >gi 3599968 (AF032123) clp protease [Arabidopsis thaliana] Length = 310
203	2023203	Tyr_Phospho_Site(964-971)
204	2023204	1E-127 >emb CAA04386 (AJ000886) Tetrafunctional protein of glyoxysomal fatty acid beta-oxidation [Brassica napus] Length = 725

205	2023205	4E-32 >emb CAA04124 (AJ000486) methionine gamma-lyase [Trichomonas vaginalis] Length = 396
206	2023206	5E-61 >pir S66770 probable membrane protein YOL077c - yeast (Saccharomyces cerevisiae) >gi 1419909 emb CAA99087 (Z74819) ORF YOL077c [Saccharomyces cerevisiae] Length = 291
207	2023207	1E-127 >emb CAA66785 (X98108) 23 kDa polypeptide of oxygen-evolving complex (OEC) [Arabidopsis thaliana] Length = 263
208	2023208	1E-131 >gb AAF00659.1 AC008153_11 (AC008153) cell division related protein [Arabidopsis thaliana] Length = 663
209	2023209	1E-141 >sp P11574 VATB_ARATH VACUOLAR ATP SYNTHASE SUBUNIT B (V-ATPASE B SUBUNIT) (V-ATPASE 57 KD SUBUNIT) >gi 81637 pir A31886 H+-transporting ATPase (EC 3.6.1.35) 57K chain - Arabidopsis thaliana >gi 166627 (J04185) nucleotide-binding subunit of vacuolar ATPase [Arabidopsis thaliana] Length = 492
210	2023210	3E-45 >gi 3242706 (AC003040) cyclin-dependent kinase inhibitor protein [Arabidopsis thaliana] >gi 3550262 (AF079587) cyclin-dependent kinase inhibitor; ICK1 [Arabidopsis thaliana] Length = 191
211	2023211	1E-140 >gb AAD28777.1 AF134130_1 (AF134130) Lhcb6 protein [Arabidopsis thaliana] Length = 258
212	2023212	1E-151) >sp P29511 TBA6_ARATH TUBULIN ALPHA-6 CHAIN >gi 282852 pir JQ1597 tubulin alpha-6 chain - Arabidopsis thaliana >gi 166920 (M84699) TUA6 [Arabidopsis thaliana] >gi 2244853 emb CAB10275.1 (Z97337) tubulin alpha-6 chain (TUA6) [Arabidopsis thaliana] Length = 450
213	2023213	Tyr Phospho Site(405-412)
214	2023214	1E-175) >emb CAB16823.1 (Z99708) aminopeptidase-like protein [Arabidopsis thaliana] Length = 634
215	2023215	2E-33 >emb CAB13047 (Z99110) yjcl [Bacillus subtilis] Length = 396
216	2023216	1E-143 >sp Q05466 HAT4_ARATH HOMEBOX-LEUCINE ZIPPER PROTEIN HAT4 (HD-ZIP PROTEIN 4) (HD-ZIP PROTEIN ATHB-2) >gi 629516 pir S31424 homeotic protein Athb-2 - Arabidopsis thaliana >gi 16180 emb CAA48246 (X68145) Athb-2 [Arabido
217	2023217	1E-149 >emb CAA72487 (Y11791) peroxidase ATP26a [Arabidopsis thaliana] Length = 276
218	2023218	Tyr Phospho Site(404-411)
219	2023219	1E-138 >gi 2262167 (AC002329) cytosolic ribosomal protein S4 [Arabidopsis thaliana] Length = 261
220	2023220	1E-163 >gb AAD30579.1 AC007260_10 (AC007260) Similar to dTDP-D-glucose 4,6-dehydratase [Arabidopsis thaliana] Length = 669
221	2023221	0) >pir S52150 serine O-acetyltransferase (EC 2.3.1.30) - Arabidopsis thaliana >gi 2146776 pir S67482 serine O-acetyltransferase (EC 2.3.1.30) - Arabidopsis thaliana >gi 608577 (L34076) serine acetyltransferase [Arabidopsis thaliana] >gi 608677 emb CAA84371 (Z348
222	2023222	1E-116 >emb CAB42903.1 (AL049862) UTP-glucose glucosyltransferase like protein [Arabidopsis thaliana] Length = 478
223	2023223	1E-46 >emb CAB10538.2 (Z97343) TEGT protein homolog [Arabidopsis thaliana] Length = 262
224	2023224	Tyr Phospho Site(1002-1010)
225	2023225	1E-117 >gi 2583121 (AC002387) phosphotransferase [Arabidopsis thaliana] Length = 257
226	2023226	Tyr Phospho Site(732-738)
227	2023227	Tyr Phospho Site(1093-1100)
228	2023228	3E-24 >gb AAD23651.1 AC007119_17 (AC007119) glycine-rich RNA binding protein Ccr2 [Arabidopsis thaliana] Length = 179
229	2023229	1E-145 >dbj BAA34250 (AB013886) RAV1 [Arabidopsis thaliana] Length = 344

230	2023230	1E-142 >emb CAB43855.1 (AL078465) isp4 like protein [Arabidopsis thaliana] Length = 753
231	2023231	4E-89 >gi 2252866 (AF013294) contains region of similarity to SYT [Arabidopsis thaliana] Length = 230
232	2023232	3E-27 >dbj BAA83740.1 (AB023288) TRAB1 [Oryza sativa] Length = 318
233	2023233	Tyr_Phospho_Site(919-926)
234	2023234	Tyr_Phospho_Site(1189-1196)
235	2023235	Tyr_Phospho_Site(301-307)
236	2023236	1E-168 >gb AAD56290.1 AF162279_1 (AF162279) 10-formyltetrahydrofolate synthetase [Arabidopsis thaliana] Length = 634
237	2023237	1E-112 >gi 3738320 (AC005170) cinnamoyl CoA reductase [Arabidopsis thaliana] Length = 303
238	2023238	1E-18 >emb CAA23041.1 (AL035394) Ap2 domain protein [Arabidopsis thaliana] Length = 343
239	2023239	Tyr_Phospho_Site(393-401)
240	2023240	4E-22 >gi 699154 (U15180) P450 cytochrome, isopentenyltransf, ferriox. [Mycobacterium leprae] Length = 187
241	2023241	1E-131 >sp P24636 TBB4_ARATH TUBULIN BETA-4 CHAIN >gi 2129546 pir S68122 beta-tubulin 4 - Arabidopsis thaliana >gi 166640 (M21415) beta-tubulin [Arabidopsis thaliana] Length = 444
242	2023242	1E-112 >gi 3790581 (AF079179) RING-H2 finger protein RHB1a [Arabidopsis thaliana] Length = 190
243	2023243	1E-124 >emb CAA55006 (X78116) Acetoacetyl-coenzyme A thiolase [Raphanus sativus] Length = 406
244	2023244	7E-11 >gi 2622337 (AE000890) inosine-5'-monophosphate dehydrogenase related protein V [Methanobacterium thermoautotrophicum] Length = 187
245	2023245	3E-11 >emb CAB45565.1 (AL079355) phospholipase C [Streptomyces coelicolor] Length = 501
246	2023246	Tyr_Phospho_Site(1121-1127)
247	2023247	1E-148 >pir S25677 chlorophyll a/b-binding protein type I precursor Lhb1B1 - Arabidopsis thaliana >gi 16366 emb CAA45789 (X64459) photosystem II type I chlorophyll a /b binding protein [Arabidopsis thaliana] >gi 3128229 (AC004077) photosystem II type I chlorophyll a/b binding protein [Arabidopsis thaliana] >gi 3337372 (AC004481) photosystem II type I chlorophyll a/b binding protein [Arabidopsis thaliana] Length = 266
248	2023248	1E-113 >gi 3941466 (AF062887) transcription factor [Arabidopsis thaliana] Length = 352
249	2023249	3E-18 >gb AAD42398.1 AF157493_6 (AF157493) carboxymethylenebutenolidase [Zymomonas mobilis] Length = 310
250	2023250	Tyr_Phospho_Site(663-671)
251	2023251	Tyr_Phospho_Site(648-655)
252	2023252	1E-138 >gb AAC62791.1 (AF096371) contains similarity to D-isomer specific 2-hydroxyacid dehydrogenases (Pfam: 2-Hacid_DH.hmm, score: 19.11) [Arabidopsis thaliana] Length = 662
253	2023253	Tyr_Phospho_Site(984-990)
254	2023254	1E-130 >sp P42737 CAH2_ARATH CARBONIC ANHYDRASE 2 (CARBONATE DEHYDRATASE 2) >gi 438449 (L18901) carbonic anhydrase [Arabidopsis thaliana] Length = 259
255	2023255	1E-135 >emb CAB39787.1 (AL049488) chlorophyll a/b-binding protein-like [Arabidopsis thaliana] >gi 4741958 gb AAD28776.1 AF134129_1 (AF134129) Lhcb5 protein [Arabidopsis thaliana] Length = 280
256	2023256	Tyr_Phospho_Site(1564-1570)
257	2023257	1E-140 >gi 3264805 (AF071788) phosphoenolpyruvate carboxylase [Arabidopsis thaliana] >gi 4079630 emb CAA10486 (AJ131710) phospho enole

		pyruvate carboxylase [Arabidopsis thaliana] Length = 968
258	2023258	1E-111 >emb CAB10530.1 (Z97343) EREBP-4 like protein [Arabidopsis thaliana] Length = 603
259	2023259	1E-127 >sp P48491 TPIS_ARATH TRIOSEPHOSPHATE ISOMERASE, CYTOSOLIC (TIM) >gi 414550 (U02949) cytosolic triose phosphate isomerase [Arabidopsis thaliana] >gi 742408 prf 2009415A triose phosphate isomerase [Arabidopsis thaliana] Length = 254
260	2023260	Tyr Phospho Site(963-969)
261	2023261	1E-152 >emb CAB36755.1 (AL035523) protein-methionine-S-oxide reductase [Arabidopsis thaliana] Length = 258
262	2023262	Tyr Phospho Site(1080-1087)
263	2023263	1E-140 >sp Q38997 K110_ARATH SNF1-RELATED PROTEIN KINASE KIN10 (AKIN10) >gi 322596 pir JC1446 serine/threonine protein kinase (EC 2.7.-.-) AK21 - Arabidopsis thaliana >gi 166600 (M93023) SNF1-related protein kinase [Arabidopsis thaliana] >gi 1742969 emb CAA64384 (X94757) ser/thr protein kinase [Arabidopsis thaliana] Length = 512
264	2023264	1E-158 >gb AAD28774.1 AF134127_1 (AF134127) Lhcb4.2 protein [Arabidopsis thaliana] Length = 287
265	2023265	Tyr Phospho Site(370-377)
266	2023266	1E-173 >gb AAD25800.1 AC006550_8 (AC006550) Identical to gb U12536 3-methylcrotonyl-CoA carboxylase precursor protein from Arabidopsis thaliana. ESTs gb H35836, gb AA651295 and gb AA721862 come from this gene. Length = 730
267	2023267	Tyr Phospho Site(861-867)
268	2023268	1E-131 >gi 3941522 (AF062915) transcription factor [Arabidopsis thaliana] Length = 249
269	2023269	1E-147 >gb AAB53256.1 (U66408) GTP-binding protein [Arabidopsis thaliana] >gi 2345150 gb AAB67830 (AF014822) developmentally regulated GTP binding protein [Arabidopsis thaliana] Length = 399
270	2023270	Tyr Phospho Site(786-793)
271	2023271	1E-133 >gi 3746809 (AF082882) adenylate kinase [Arabidopsis thaliana] Length = 246
272	2023272	3E-91 >emb CAA71277 (Y10228) P-glycoprotein-2 [Arabidopsis thaliana] >gi 2108254 emb CAA71276 (Y10227) P-glycoprotein-2 [Arabidopsis thaliana] >gi 4538925 emb CAB39661.1 (AL049483) P-glycoprotein-2 (pgp2) [Arabidopsis thaliana] Length = 1233
273	2023273	1E-107 >gi 1353352 (U31975) alanine aminotransferase [Chlamydomonas reinhardtii] Length = 521
274	2023274	7E-84 >emb CAA23040.1 (AL035394) receptor kinase [Arabidopsis thaliana] Length = 638
275	2023275	1E-129 >gi 1145697 (U39485) delta tonoplast integral protein [Arabidopsis thaliana] Length = 250
276	2023276	1E-54 >emb CAA96657.1 (Z72511) possible zinc finger protein; cDNA EST EMBL:M89115 comes from this gene; cDNA EST EMBL:D71533 comes from this gene; cDNA EST EMBL:D72314 comes from this gene; cDNA EST EMBL:D75164 comes from this gene; cDNA EST EMBL:... Length = 610
277	2023277	Pkc Phospho Site(73-75)
278	2023278	1E-154 >gi 3335374 (AC003028) glutaredoxin-like protein [Arabidopsis thaliana] Length = 293
279	2023279	1E-128 >gb AAD57005.1 AC009465_19 (AC009465) 40S ribosomal protein S3A (S phase specific) [Arabidopsis thaliana] Length = 262
280	2023280	1E-114 >gb AAD28778.1 AF134131_1 (AF134131) PsbS protein [Arabidopsis thaliana] Length = 265
281	2023281	7E-62 >gb AAD25756.1 AC007060_14 (AC007060) Contains the PF 00650 CRAL/TRIO phosphatidyl-inositol-transfer protein domain. ESTs gb T76582,

		gb N06574 and gb Z25700 come from this gene. [Arabidopsis thaliana] Length = 540
282	2023282	0 >sp P25851 F16P_ARATH FRUCTOSE-1,6-BISPHOSPHATASE, CHLOROPLAST PRECURSOR (D-FRUCTOSE-1,6-BISPHOSPHATE 1-PHOSPHOHYDROLASE) (FBPASE) >gi 99693 pir S16582 fructose-bisphosphatase (EC 3.1.3.11) precursor, chloroplast - Arabidopsis thaliana >gi 11242 emb CAA41154 (X58148) fructose-bisphosphatase [Arabidopsis thaliana] Length = 417
283	2023283	1E-162 >gi 4220476 (AC006069) ribophorin I-like protein [Arabidopsis thaliana] Length = 464
284	2023284	1E-151 >pir UQPM ubiquitin precursor - garden pea >gi 20589 emb CAA34886 (X17020) polyubiquitin (AA 1-381) [Pisum sativum] >gi 4115339 (L81142) ubiquitin [Pisum sativum] >gi 226707 prf 1603402A polyubiquitin [Pisum sativum] Length = 381
285	2023285	Rgd(1319-1321)
286	2023286	1E-143 >gi 3980379 (AC004561) cyclin, PCNA [Arabidopsis thaliana] Length = 264
287	2023287	1E-108 >gb AAF00071.1 AF093604_1 (AF093604) apyrase [Arabidopsis thaliana] Length = 471
288	2023288	8E-99 >sp P36397 ARF1_ARATH ADP-RIBOSYLATION FACTOR 1 >gi 322518 pir S28875 ADP-ribosylation factor 1 - Arabidopsis thaliana >gi 166586 (M95166) ADP-ribosylation factor [Arabidopsis thaliana] >gi 2275195 (AC002337) ADP-ribosylation factor [Arabidopsis thaliana] >gi 4630747 gb AAD26597.1 AC007236_2 (AC007236) ADP-ribosylation factor [Arabidopsis thaliana] Length = 181
289	2023289	Tyr Phospho Site(570-577)
290	2023290	Zinc Finger_C3hc4(177-186)
291	2023291	Pkc Phospho Site(23-25)
292	2023292	1E-146 >emb CAB43632.1 (AL050351) SEC14-like protein [Arabidopsis thaliana] Length = 617
293	2023293	1E-109 >sp P46422 GTH4_ARATH GLUTATHIONE S-TRANSFERASE PM24 (24 KD AUXIN-BINDING PROTEIN) (GST CLASS PHI) >gi 479736 pir S35268 glutathione transferase (EC 2.5.1.18) gst2 - Arabidopsis thaliana >gi 166723 (L07589) glutathione S-transferase [Arabidopsis thaliana] >gi 347212 (L11601) glutathione S-transferase [Arabidopsis thaliana] >gi 407090 emb CAA53051 (X75303) glutathione S-transferase [Arabidopsis thaliana] >gi 2262152 gb AAC78264.1 AAC78264 (AC002330) Atpm24.1 glutathione S transferase [Arabidopsis thaliana] Length = 212
294	2023294	3E-21 >emb CAA22977.1 (AL035353) photosystem I subunit PSI-E-like protein [Arabidopsis thaliana] >gi 5732203 emb CAB52678.1 (AJ245908) photosystem I subunit IV precursor [Arabidopsis thaliana] Length = 143
295	2023295	Tyr Phospho Site(441-447)
296	2023296	1E-159 >gi 166835 (M86720) ribulose bisphosphate carboxylase/oxygenase activase [Arabidopsis thaliana] >gi 2642170 (AC003000) Rubisco activase [Arabidopsis thaliana] Length = 446
297	2023297	Tyr Phospho Site(757-764)
298	2023298	1E-22 >gi 4102690 (AF004806) 24 kDa seed maturation protein [Glycine max] Length = 212
299	2023299	Tyr Phospho Site(366-373)
300	2023300	1E-142 >gi 4056500 (AC005896) acetyltransferase [Arabidopsis thaliana] Length = 432
301	2023301	5E-68 >emb CAA07236 (AJ006771) beta-galactosidase [Cicer arietinum] Length = 707
302	2023302	1E-104 >sp P52577 IFRH_ARATH ISOFLAVONE REDUCTASE HOMOLOG P3 >gi 1361992 pir S57613 isoflavonoid reductase homolog - Arabidopsis thaliana

		>gi 886432 emb CAA89859 (Z49777) isoflavonoid reductase homologue [Arabidopsis thaliana]
303	2023303	1E-123 >gb AAD20405 (AC007019) ATP synthase [Arabidopsis thaliana] Length = 240
304	2023304	1E-131 >dbj BAA32418 (AB008103) ethylene responsive element binding factor 1 [Arabidopsis thaliana] Length = 266
305	2023305	1E-142 >dbj BAA78560.1 (AB024282) cysteine synthase [Arabidopsis thaliana] >gi 5824334 emb CAB54830.1 (AJ010505) cysteine synthase [Arabidopsis thaliana] Length = 368
306	2023306	Tyr Phospho Site(92-100)
307	2023307	2E-79 >emb CAB42925.1 (AL049862) tRNA synthetase [Arabidopsis thaliana] Length = 225
308	2023308	3E-25 >gb AAD46141.1 AF081022_1 (AF081022) hypoxia-induced protein L31 [Lycopersicon esculentum] Length = 78
309	2023309	1E-110 >emb CAA16677 (AL021684) LRR-like protein [Arabidopsis thaliana] Length = 445
310	2023310	8E-38 >dbj BAA22374 (D86122) Mei2-like protein [Arabidopsis thaliana] Length = 884
311	2023311	1E-135 >gb AAD32291.1 AC006533_15 (AC006533) acetolactate synthase [Arabidopsis thaliana] Length = 484
312	2023312	2E-98 >gb AAB51567.1 (U75189) germin-like protein [Arabidopsis thaliana] >gi 1755158 gb AAB51568.1 (U75190) germin-like protein [Arabidopsis thaliana] >gi 1755170 gb AAB51574.1 (U75196) germin-like protein [Arabidopsis thaliana] >gi 1755172 gb AAB51575.1 (U75197) germin-like protein [Arabidopsis thaliana] >gi 1755180 gb AAB51579.1 (U75201) germin-like protein [Arabidopsis thaliana] >gi 1755190 gb AAB51584.1 (U75206) germin-like protein [Arabidopsis thaliana] >gi 1934728 gb AAB51751.1 (U95035) germin-like protein [Arabidopsis thaliana] >gi 4154285 (AF090733) germin-like protein 1 [Arabidopsis thaliana] >gi 4666248 dbj BAA77207.1 (D89055) germin-like protein precursor [Arabidopsis thaliana] Length = 208
313	2023313	Pkc Phospho Site(14-16)
314	2023314	Pkc Phospho Site(92-94)
315	2023315	1E-119 >emb CAA96434 (Z71752) pectin methylesterase [Nicotiana glauca] Length = 315
316	2023316	1E-130) >sp O23708 PRC3_ARATH PROTEASOME COMPONENT C3 (MACROPAIN SUBUNIT C3) (MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT C3) >gi 2511574 emb CAA73619.1 (Y13176) multicatalytic endopeptidase [Arabidopsis thaliana] >gi 3421075 (AF043520) 20S proteasome subunit PAB1 [Arabidopsis thaliana] >gi 4966368 gb AAD34699.1 AC006341_27 (AC006341) Identical to gb Y13176 Arabidopsis thaliana mRNA for proteasome subunit prc3. ESTs gb H36972, gb T22551 and gb T13800 come from this gene. Length = 235
317	2023317	Pkc Phospho Site(11-13)
318	2023318	Tyr Phospho Site(1345-1353)
319	2023319	Tyr Phospho Site(309-315)
320	2023320	1E-115 >gi 2829275 (AF044265) nucleoside diphosphate kinase 3 [Arabidopsis thaliana] >gi 3513740 (AF080118) contains similarity to nucleoside diphosphate kinases (Pfam: NDK.hmm, score: 301.12) [Arabidopsis thaliana] >gi 4539375 emb CAB40069.1 (AL049525) nucleoside diphosphate kinase 3 (ndpk3) [Arabidopsis thaliana] Length = 238
321	2023321	1E-160 >sp P42498 PHYE_ARATH PHYTOCHROME E >gi 1076376 pir S46313 phytochrome E - Arabidopsis thaliana >gi 452817 emb CAA54075 (X76610) phytochrome E [Arabidopsis thaliana] >gi 5816999 emb CAB53654.1 (AL110123) phytochrome E [Arabidopsis thaliana] Length = 1112

322	2023322	1E-35 >gb AAD28506.1 AF123265_1 (AF123265) remorin 1 [Lycopersicon esculentum] Length = 197
323	2023323	1E-171 >gi 4220452 (AC006216) Similar to gi 3413714 T19L18.21 myrosinase-binding protein from Arabidopsis thaliana BAC gb AC004747. ESTs gb T44298, gb T42447, gb R64761 and gb 100206 come from this gene. [Arabidopsis thaliana] Length = 292
324	2023324	3E-21 >pir S62011 PHO85 protein - yeast (Saccharomyces cerevisiae) >gi 1163103 (U43503) Lph16p [Saccharomyces cerevisiae] Length = 1223
325	2023325	4E-59 >sp P73839 THDF_SYNY3 POSSIBLE THIOPHENE AND FURAN OXIDATION PROTEIN THDF >gi 1652979 dbj BAA17896 (D90910) thiophen and furan oxidation protein [Synechocystis sp.] Length = 456
326	2023326	1E-117 >emb CAA17161 (AL021890) calcium-dependent protein kinase - like protein [Arabidopsis thaliana] >gi 2961339 emb CAA18097.1 (AL022140) calcium-dependent protein kinase-like protein [Arabidopsis thaliana] Length = 554
327	2023327	1E-105 >gi 3980412 (AC004561) pumilio-like protein [Arabidopsis thaliana] Length = 968
328	2023328	1E-160 >dbj BAA82066.1 (AB022327) nClpP2 [Arabidopsis thaliana] Length = 279
329	2023329	1E-129 >emb CAA04172 (AJ000539) phosphatidylinositol synthase [Arabidopsis thaliana] Length = 227
330	2023330	8E-65 >gb AAD11598.1 AAD11598 (AF071527) calcium channel [Arabidopsis thaliana] >gi 4263043 gb AAD15312 (AC005142) calcium channel [Arabidopsis thaliana] Length = 724
331	2023331	Tyr Phospho Site(46-53)
332	2023332	1E-126 >gi 2981475 (AF053084) cinnamyl alcohol dehydrogenase [Malus domestica] Length = 325
333	2023333	Tyr Phospho Site(126-132)
334	2023334	1E-142 >emb CAB39936.1 (AL049500) osmotin precursor [Arabidopsis thaliana] Length = 244
335	2023335	1E-138 >gb AAD28767.1 AF134120_1 (AF134120) Lhca2 protein [Arabidopsis thaliana] Length = 257
336	2023336	Tyr Phospho Site(628-636)
337	2023337	3E-14 >sp P34092 MYSB_DICDI MYOSIN IB HEAVY CHAIN >gi 102252 pir A33284 myosin heavy chain IB - slime mold (Dictyostelium discoideum) >gi 167839 (M26037) myosin I heavy chain [Dictyostelium discoideum] Length = 1111
338	2023338	2E-68 >sp P37707 B2_DAUCA B2 PROTEIN >gi 322726 pir S32124 B2 protein - carrot >gi 297889 emb CAA51078 (X72385) B2 protein [Daucus carota] Length = 207
339	2023339	1E-146 >gi 3980402 (AC004561) tropinone reductase [Arabidopsis thaliana] Length = 260
340	2023340	1E-68 >dbj BAA11226 (D78151) human 26S proteasome subunit p97 [Homo sapiens] Length = 908
341	2023341	1E-117 >sp P51430 RS6_ARATH 40S RIBOSOMAL PROTEIN S6 >gi 2224751 emb CAA74381 (Y14052) ribosomal protein S6 [Arabidopsis thaliana] Length = 249
342	2023342	1E-109 >emb CAA17550 (AL021961) receptor protein kinase - like protein [Arabidopsis thaliana] Length = 980
343	2023343	1E-106 >sp Q42599 NUIM_ARATH NADH-UBIQUINONE OXIDOREDUCTASE 23 KD SUBUNIT PRECURSOR (COMPLEX I-23KD) (CI- 23KD) >gi 1076356 pir S52380 NADH dehydrogenase (EC 1.6.99.3) - Arabidopsis thaliana >gi 666977 emb CAA59061 (X84318) NADH dehydrogenase [Arabidopsis thaliana] >gi 3152573 (AC002986) Match to NADH:ubiquinone oxidoreductase gb X84318 from A.thaliana. ESTs gb Z27005, gb T04711, gb T45078 and gb Z28689 come from this gene. [Arabidopsis thaliana] Length =

		222
344	2023344	1E-142) >gi 3763918 (AC004450) isopropylmalate dehydratase [Arabidopsis thaliana] Length = 251
345	2023345	5E-84 >sp P54641 VATX_DICDI VACUOLAR ATP SYNTHASE SUBUNIT AC39 (V-ATPASE AC39 SUBUNIT) (41 KD ACCESSORY PROTEIN) (DVA41) >gi 626048 pir A55016 lysosomal membrane protein DVA41 - slime mold (Dictyostelium discoideum) >gi 532733 (U13150) vacuolar ATPase subunit DVA41 [Dictyostelium discoideum] Length = 356
346	2023346	5E-88 >gb AAD15451 (AC006068) receptor protein kinase [Arabidopsis thaliana] Length = 567
347	2023347	1E-61 >sp P31166 APT1_ARATH ADENINE PHOSPHORIBOSYLTRANSFERASE 1 (APRT) >gi 99657 pir S20867 adenine phosphoribosyltransferase (EC 2.4.2.7) - Arabidopsis thaliana >gi 16164 emb CAA41497 (X58640) adenine phosphoribosyltransferase [Arabidopsis thaliana] >gi 433050 (L19637) adenine phosphoribosyltransferase [Arabidopsis thaliana] >gi 3935182 (AC004557) F17L21.25 [Arabidopsis thaliana] Length = 183
348	2023348	1E-127 >emb CAA10060.1 (AJ012571) glutathione transferase [Arabidopsis thaliana] Length = 219
349	2023349	Pkc_Phospho_Site(28-30)
350	2023350	1E-123 >gi 3201613 (AC004669) glutathione S-transferase [Arabidopsis thaliana] Length = 215
351	2023351	1E-109 >sp P51119 GLN2_VITVI GLUTAMINE SYNTHETASE CYTOSOLIC ISOZYME 2 (GLUTAMATE—AMMONIA LIGASE) >gi 1134898 emb CAA63982 (X94321) glutamine synthetase [Vitis vinifera] Length = 356
352	2023352	2E-23 >gi 871782 (L43081) pEARLI 4 gene product [Arabidopsis thaliana] Length = 766
353	2023353	1E-150 >emb CAA66963 (X98319) peroxidase [Arabidopsis thaliana] >gi 1429217 emb CAA67311 (X98775) peroxidase ATP12a [Arabidopsis thaliana] Length = 321
354	2023354	8E-46 >gi 4206763 (AF104328) cell wall-plasma membrane linker protein homolog [Arabidopsis thaliana] Length = 306
355	2023355	1E-140 >gi 1644427 (U74610) glyoxalase II [Arabidopsis thaliana] Length = 256
356	2023356	1E-158 >gi 3757514 (AC005167) plasma membrane intrinsic protein [Arabidopsis thaliana] >gi 4581129 gb AAD24619.1 AC005825_26 (AC005825) plasma membrane intrinsic protein [Arabidopsis thaliana] Length = 278
357	2023357	1E-139 >gi 2708750 (AC003952) physical impedance protein [Arabidopsis thaliana] Length = 452
358	2023358	1E-117 >sp O04157 RAB7_ARATH RAS-RELATED PROTEIN RAB7 >gi 2065015 emb CAA70951 (Y09821) GTP-binding protein Rab7 [Arabidopsis thaliana] >gi 2505866 emb CAA72904 (Y12227) GTP-binding protein Rab7 [Arabidopsis thaliana] >gi 3287684 (AC003979) Strong similarity to gb Y09821 GTP-binding protein Rab7 from A. thaliana. EST gb T76449 comes from this gene. [Arabidopsis thaliana] Length = 203
359	2023359	3E-20 >gi 3213227 (AF035209) v-SNARE Vti1a [Mus musculus] >gi 3421062 (AF035823) 29-kDa Golgi SNARE [Mus musculus] Length = 217
360	2023360	2E-25 >dbj BAA37095.1 (AB022209) ribonucleoprotein F [Rattus norvegicus] Length = 415
361	2023361	Pkc_Phospho_Site(67-69)
362	2023362	6E-78 >gb AAD25780.1 AC006577_16 (AC006577) Similar to gb U55861 RNA binding protein nucleolysin (TIAR) from Mus musculus and contains several PF 00076 RNA recognition motif domains. ESTs gb T21032 and gb T44127 come from this gene. [Arabidopsis t... Length = 426
363	2023363	Pkc_Phospho_Site(14-16)

364	2023364	3E-11 >emb CAA16558 (AL021635) leucine rich repeat receptor kinase-like protein [Arabidopsis thaliana] Length = 688
365	2023365	1E-140 >sp P34791 CYP4_ARATH PEPTIDYL-PROLYL CIS-TRANS ISOMERASE, CHLOROPLAST PRECURSOR (PPIASE) (ROTAMASE) (CYCLOPHILIN) (CYCLOSPORIN A-BINDING PROTEIN) >gi 1076368 pir B53422 peptidylprolyl isomerase (EC 5.2.1.8) ROC4 - Arabidopsis thaliana >gi 405131 (L14845) cyclophilin [Arabidopsis thaliana] >gi 1322278 (U42724) cyclophilin [Arabidopsis thaliana] Length = 260
366	2023366	2E-56 >emb CAA89697 (Z49697) cysteine proteinase inhibitor [Ricinus communis] Length = 209
367	2023367	Tyr Phospho Site(1552-1558)
368	2023368	1E-137 >gi 2252855 (AF013294) similar to the myc family of helix-loop-helix transcription factors [Arabidopsis thaliana] Length = 423
369	2023369	1E-103 >sp P48006 EF1B_ARATH ELONGATION FACTOR 1-BETA A1 (EF-1-BETA) >gi 480620 pir S37103 translation elongation factor eEF-1 beta-A1 chain - Arabidopsis thaliana (cv. Colombia) >gi 398608 emb CAA52751 (X74733) elongation factor-1 beta A1 [Arabidopsis thaliana] Length = 231
370	2023370	1E-109 >emb CAA74639 (Y14251) glutathione S-transferase [Arabidopsis thaliana] Length = 209
371	2023371	Rgd(581-583)
372	2023372	1E-131 >gb AAD51783.1 AF145300_1 (AF145300) 14-3-3 protein GF14 kappa [Arabidopsis thaliana] Length = 248
373	2023373	1E-139 >emb CAA51171 (X72581) tonoplast intrinsic protein gamma (gamma-TIP) [Arabidopsis thaliana] Length = 251
374	2023374	Tyr Phospho Site(1037-1044)
375	2023375	1E-126 >emb CAB10400.1 (Z97340) enoyl-CoA hydratase like protein [Arabidopsis thaliana] Length = 244
376	2023376	3E-15 >gb AAD34107.1 AF151870_1 (AF151870) CGI-112 protein [Homo sapiens] Length = 208
377	2023377	1E-137 >gb AAD25640.1 AC007170_2 (AC007170) cytoplasmic aconitate hydratase [Arabidopsis thaliana] Length = 898
378	2023378	Tyr Phospho Site(787-793)
379	2023379	1E-123 >sp P52032 GSHY_ARATH GLUTATHIONE PEROXIDASE HOMOLOG PRECURSOR >gi 2129599 pir S71250 glutathione peroxidase - Arabidopsis thaliana >gi 1061036 emb CAA61965 (X89866) glutathione peroxidase [Arabidopsis thaliana] Length = 242
380	2023380	3E-99 >gb AAD25928.1 AF085279_1 (AF085279) hypothetical Ser-Thr protein kinase [Arabidopsis thaliana] Length = 570
381	2023381	6E-58 >emb CAB43976.1 (AL078579) zinc finger protein [Arabidopsis thaliana] Length = 327
382	2023382	1E-132 >gi 3421087 (AF043524) 20S proteasome subunit PAE1 [Arabidopsis thaliana] >gi 6056394 gb AAF02858.1 AC009324_7 (AC009324) 20S proteasome subunit PAE1 [Arabidopsis thaliana] Length = 237
383	2023383	2E-14 >emb CAA92677.1 (Z68315) Similarity to Human MAP kinase phosphatase-1 (SW:PTN7 HUMAN) [Caenorhabditis elegans] Length = 150
384	2023384	1E-146 >gb AAD37165.1 AF132742_1 (AF132742) 3-phosphoinositide-dependent protein kinase-1 [Arabidopsis thaliana] Length = 491
385	2023385	1E-109 >emb CAA64820 (X95573) salt-tolerance zinc finger protein [Arabidopsis thaliana] Length = 227
386	2023386	1E-169 >gi 3834309 (AC005679) Strong similarity to glycoprotein EP1 gb L16983 Daucus carota and a member of S locus glycoprotein family PF 00954. ESTs gb F13813, gb T21052, gb R30218 and gb W43262 come from this gene.
387	2023387	4E-20 >ref NP_006283.1 PTSG101 tumor susceptibility gene 101 >gi 3184258 (U82130) tumor susceptibility protein [Homo sapiens] Length = 390
388	2023388	1E-163 >gi 1046225 (U21952) ethylene response sensor [Arabidopsis

		thaliana] >gi 2623308 (AC002409) ethylene response sensor (ERS) [Arabidopsis thaliana] >gi 1584365 prf 2122405A ERS gene [Arabidopsis thaliana] Length = 613
389	2023389	Tyr Phospho Site(86-93)
390	2023390	1E-138 >sp Q08733 WC1C_ARATH PLASMA MEMBRANE INTRINSIC PROTEIN 1C (TRANSMEMBRANE PROTEIN B) (TMP-B) >gi 396218 emb CAA49155 (X69294) transmembrane protein TMP-B [Arabidopsis thaliana] Length = 286
391	2023391	7E-28 >dbj BAA32422 (AB008107) ethylene responsive element binding factor 5 [Arabidopsis thaliana] Length = 300
392	2023392	1E-108 >dbj BAA31509 (AB010877) chloroplast ribosomal protein L3 [Nicotiana tabacum] Length = 259
393	2023393	Pkc Phospho Site(133-135)
394	2023394	Tyr Phospho Site(1037-1043)
395	2023395	Tyr Phospho Site(603-609)
396	2023396	Tyr Phospho Site(579-586)
397	2023397	1E-101 >dbj BAA25180 (D88536) delta 9 desaturase [Arabidopsis thaliana] Length = 305
398	2023398	Tyr Phospho Site(1372-1378)
399	2023399	1E-105 >emb CAB08077 (Z94058) pectinesterase [Lycopersicon esculentum] Length = 504
400	2023400	4E-35 >emb CAA19765 (AL031004) RSZp22 splicing factor [Arabidopsis thaliana] >gi 3435094 gb AAD12769.1 (AF033586) 9G8-like SR protein [Arabidopsis thaliana] Length = 200
401	2023401	1E-125 >gi 2191150 (AF007269) similar to mitochondrial carrier family [Arabidopsis thaliana] Length = 352
402	2023402	1E-136 >emb CAA74025.1 (Y13691) multicatalytic endopeptidase complex, proteasome component, alpha subunit [Arabidopsis thaliana] Length = 245
403	2023403	1E-156 >sp P25697 KPPR_ARATH PHOSPHORIBULOKINASE PRECURSOR (PHOSPHOPENTOKINASE) (PRKASE) (PRK) >gi 99744 pir S16583 phosphoribulokinase (EC 2.7.1.19) precursor - Arabidopsis thaliana >gi 16441 emb CAA41155 (X58149) Ribulose-5-phosphate kinase [Arabidopsis thaliana] Length = 395
404	2023404	1E-90 >dbj BAA77837.1 (AB027458) ACE [Arabidopsis thaliana] >gi 5903086 gb AAD55644.1 AC008017_17 (AC008017) ACE [Arabidopsis thaliana] Length = 594
405	2023405	1E-98 >dbj BAA24804 (AB010946) AtRer1B [Arabidopsis thaliana] Length = 195
406	2023406	Tyr Phospho Site(120-126)
407	2023407	1E-143 >gb AAD39331.1 AC007258_20 (AC007258) pyruvate dehydrogenase E1 alpha subunit [Arabidopsis thaliana] Length = 389
408	2023408	Tyr Phospho Site(593-601)
409	2023409	1E-14 >gi 3152583 (AC002986) Contains similarity to inhibitor of apoptosis protein gb U45881 from D. melanogaster. [Arabidopsis thaliana] Length = 347
410	2023410	Tyr Phospho Site(1596-1603)
411	2023411	Tyr Phospho Site(1068-1075)
412	2023412	1E-127 >gb AAD31074.1 AC007357_23 (AC007357) Similar to gb AF038007 FIC1 gene from Homo sapiens and is a member of the PF 00122 E1-E2 ATPase family. ESTs gb T45045 and gb AA394473 come from this gene. [Arabidopsis thaliana] Length = 1203
413	2023413	1E-123 >gi 2583123 (AC002387) nucleotide sugar epimerase [Arabidopsis thaliana] Length = 437
414	2023414	1E-127 >gb AAD28780.1 AF134133_1 (AF134133) Lil3 protein [Arabidopsis

		thaliana] Length = 262
415	2023415	3E-94 >gi 2511546 (AF022658) c2h2 zinc finger transcription factor [Arabidopsis thaliana] Length = 238
416	2023416	Tyr Phospho Site(724-732)
417	2023417	1E-123 >gi 2618723 (U49073) IAA17 [Arabidopsis thaliana] >gi 2921756 (AF040631) IAA17/AXR3 protein [Arabidopsis thaliana] >gi 4389514 gb AAB70451 (AC000104) Identical to Arabidopsis gb AF040632 and gb U49073 IAA17/AXR3 gene. ESTs gb H36782 and gb F14074 come from this gene. [Arabidopsis thaliana] Length = 229
418	2023418	1E-157 >gi 4138855 (AF098072) IMMUTANS [Arabidopsis thaliana] Length = 351
419	2023419	Tyr Phospho Site(1298-1305)
420	2023420	3E-41 >gb AAD45585.1 AF132115_1 (AF132115) cytochrome b-561 [Arabidopsis thaliana] Length = 230
421	2023421	1E-127 >pir S25435 chlorophyll a/b-binding protein - Arabidopsis thaliana >gi 16207 emb CAA39534 (X56062) chlorophyll A/B-binding protein [Arabidopsis thaliana] >gi 166644 (M85150) chlorophyll a/b-binding protein [Arabidopsis thaliana] >gi 4678304 emb CAB41095.1 (AL049655) chlorophyll a/b-binding protein [Arabidopsis thaliana] Length = 241
422	2023422	1E-148 >sp P21216 IPYR_ARATH SOLUBLE INORGANIC PYROPHOSPHATASE (PYROPHOSPHATE PHOSPHO-HYDROLASE) (PPASE) >gi 81645 pir S13379 inorganic pyrophosphatase (EC 3.6.1.1) - Arabidopsis thaliana >gi 16348 emb CAA40764 (X57545) inorganic pyrophosphatase [Arabidopsis thaliana] Length = 263
423	2023423	8E-69 >gi 3928094 (AC005770) zinc finger protein [Arabidopsis thaliana] Length = 270
424	2023424	2E-57 >emb CAA77089 (Y18227) blue copper binding-like protein [Arabidopsis thaliana] Length = 196
425	2023425	1E-149 >emb CAA18252.1 (AL022224) CLV1 receptor kinase like protein [Arabidopsis thaliana] Length = 992
426	2023426	Tyr Phospho Site(935-942)
427	2023427	1E-157 >gb AAD18142 (AC006260) plasma membrane intrinsic protein 2B [Arabidopsis thaliana] Length = 285
428	2023428	Tyr Phospho Site(699-707)
429	2023429	1E-125) >gb AAD24640.1 AC006919_18 (AC006919) pyruvate kinase [Arabidopsis thaliana] Length = 464
430	2023430	Rgd(1781-1783)
431	2023431	1E-134 >gb AAD24630.1 AC006919_8 (AC006919) fructose-bisphosphate aldolase, cytoplasmic [Arabidopsis thaliana] Length = 358
432	2023432	Pkc Phospho Site(101-103)
433	2023433	1E-136 >gi 3004557 (AC003673) plasma membrane proton pump H ⁺ -ATPase, PMA1 [Arabidopsis thaliana] Length = 949
434	2023434	1E-138) >gi 2191128 (AF007269) belongs to the L5P family of ribosomal proteins [Arabidopsis thaliana] Length = 262
435	2023435	3E-98 >gi 1946371 (U93215) regulatory protein Viviparous-1 isolog [Arabidopsis thaliana] Length = 780
436	2023436	1E-156 >gb AAD28773.1 AF134126_1 (AF134126) Lhcb3 protein [Arabidopsis thaliana] >gi 5002210 gb AAD37362.1 AF143691_1 (AF143691) type III chlorophyll a/b binding protein [Arabidopsis thaliana] Length = 265
437	2023437	7E-67 >gi 2459430 (AC002332) CUC2 protein [Arabidopsis thaliana] Length = 268
438	2023438	1E-155 >sp P04777 CB21_ARATH CHLOROPHYLL A-B BINDING PROTEIN 165/180 PRECURSOR (LHCII TYPE I CAB-165/180) (LHCP) >gi 81603 pir A29280 chlorophyll a/b-binding protein ab165 - Arabidopsis thaliana >gi 16368 emb CAA27540 (X03907) chlorophyll a/b binding protein (LHCP AB 65)

		[Arabidopsis thaliana] >gi 16372 emb CAA27541 (X03908) chlorophyll a/b binding protein (LHCP AB 180) [Arabidopsis thaliana] Length = 267
439	2023439	2E-58 >emb CAA63223 (X92491) TOM20 [Solanum tuberosum] Length = 204
440	2023440	1E-89 >emb CAB40742.1 (AJ237751) aquaglyceroporin [Nicotiana tabacum] Length = 247
441	2023441	1E-29 >gb AAD15610 (AC006232) selenium-binding protein [Arabidopsis thaliana] Length = 472
442	2023442	1E-146 >gb AAD20124 (AC006201) 60S ribosomal protein L2 [Arabidopsis thaliana] Length = 258
443	2023443	1E-125 >emb CAB45800.1 (AL080252) nodulin-like protein [Arabidopsis thaliana] Length = 368
444	2023444	Tyr Phospho Site(880-887)
445	2023445	Tyr Phospho Site(747-754)
446	2023446	Tyr Phospho Site(353-361)
447	2023447	4E-34 >gi 3421373 (AF079901) 28 kDa cis-Golgi SNARE [Mus musculus] Length = 250
448	2023448	1E-64 >sp Q43794 SYE TOBAC GLUTAMYL-TRNA SYNTHETASE (GLUTAMATE—TRNA LIGASE) (GLURS) >gi 1084418 pir S51685 glutamate—tRNA ligase (EC 6.1.1.17) - common tobacco >gi 603867 emb CAA58506 (X83524) glutamate—tRNA ligase [Nicotiana tabacum] Length = 569
449	2023449	1E-110 >emb CAB16805.1 (Z99708) minor allergen [Arabidopsis thaliana] Length = 273
450	2023450	6E-17 >gb AAD25848.1 AC007197_1 (AC007197) disease resistance gene, 5' partial [Arabidopsis thaliana] Length = 554
451	2023451	1E-65 >emb CAA74639 (Y14251) glutathione S-transferase [Arabidopsis thaliana] Length = 209
452	2023452	2E-83 >gi 2598932 (AF027157) auxin-responsive protein IAA2 [Arabidopsis thaliana] Length = 174
453	2023453	8E-56 >gi 3287683 (AC003979) Similar to apoptosis protein MA-3 gb D50465 from Mus musculus. [Arabidopsis thaliana] Length = 693
454	2023454	1E-125 >gi 1764100 (U81805) GDP-D-mannose-4,6-dehydratase [Arabidopsis thaliana] Length = 373
455	2023455	1E-109 >gi 3510259 (AC005310) inorganic pyrophosphatase [Arabidopsis thaliana] >gi 3522960 gb AAC34242.1 (AC004411) inorganic pyrophosphatase [Arabidopsis thaliana] Length = 216
456	2023456	2E-20 >emb CAA07361.1 (AJ006972) TOM1 [Mus musculus] Length = 492
457	2023457	1E-143 >gb AAD25595.1 AC007211_17 (AC007211) chlorophyll A/B binding protein [Arabidopsis thaliana] >gi 4741946 gb AAD28770.1 AF134123_1 (AF134123) Lhcb2 protein [Arabidopsis thaliana] Length = 265
458	2023458	1E-79 >gb AAD31350.1 AC007212_6 (AC007212) bZIP transcription factor [Arabidopsis thaliana] Length = 171
459	2023459	Pkc Phospho Site(2-4)
460	2023460	Pkc Phospho Site(9-11)
461	2023461	1E-146 >gi 3980396 (AC004561) C-4 sterol methyl oxidase [Arabidopsis thaliana] Length = 253
462	2023462	Tyr Phospho Site(620-626)
463	2023463	6E-81 >gi 3831468 (AC005700) phosphocholine cytidyltransferase [Arabidopsis thaliana] >gi 5640001 gb AAD45922.1 AF165912_1 (AF165912) CTP:phosphocholine cytidyltransferase [Arabidopsis thaliana] Length = 332
464	2023464	1E-153 >gi 3850579 (AC005278) Strong similarity to gb D14550 extracellular dermal glycoprotein (EDGP) precursor from Daucus carota. ESTs gb H37281, gb T44167, gb T21813, gb N38437, gb Z26470, gb R65072, gb N76373, gb F15470, gb Z35182, gb H76373, gb Z34678 an... Length = 433

465	2023465	1E-40 >sp P48724 IF5_PHAVU EUKARYOTIC TRANSLATION INITIATION FACTOR 5 (EIF-5) >gi 1008881 (L47221) eukaryotic initiation factor 5 [Phaseolus vulgaris] Length = 443
466	2023466	2E-96 >sp P42043 HMZ1_ARATH FERROCHELATASE I, CHLOROPLAST/MITOCHONDRIAL PRECURSOR (PROTOHEME FERROLYASE) (HEME SYNTHETASE) >gi 1076325 pir A54125 ferrochelatase (EC 4.99.1.1) precursor, chloroplast - Arabidopsis thaliana >gi 511081 emb CAA51819 (X73417) ferrochelatase [Arabid
467	2023467	Pkc Phospho_Site(8-10)
468	2023468	1E-132 >dbj BAA31525 (AB013301) ethylene responsive element binding factor [Arabidopsis thaliana] Length = 281
469	2023469	1E-112) >sp P28187 ARA4_ARATH RAS-RELATED PROTEIN ARA-4 >gi 81633 pir JS0641 GTP-binding protein ara-4 - Arabidopsis thaliana >gi 217839 dbj BAA00831 (D01026) small GTP-binding protein [Arabidopsis thaliana] >gi 3763922 (AC004450) GTP-binding protein [Arabidopsis thaliana] Length = 214
470	2023470	Rgd(476-478)
471	2023471	Zinc Finger C2h2(514-536)
472	2023472	2E-92 >gi 1872521 (U87833) zinc-finger protein Lsd1 [Arabidopsis thaliana] >gi 1872523 (U87834) zinc-finger protein Lsd1 [Arabidopsis thaliana] >gi 5262161 emb CAB45804.1 (AL080253) zinc-finger protein Lsd1 [Arabidopsis thaliana] Length = 189
473	2023473	1E-133 >emb CAB42872.1 (AJ012423) wall-associated kinase 2 [Arabidopsis thaliana] Length = 732
474	2023474	2E-30 >gi 2224911 (U93048) somatic embryogenesis receptor-like kinase [Daucus carota] Length = 553
475	2023475	Tyr Phospho_Site(869-875)
476	2023476	3E-46 >dbj BAA25999 (AB013447) aluminum-induced [Brassica napus] Length = 244
477	2023477	Rgd(263-265)
478	2023478	1E-104) >emb CAA70498 (Y09314) Rab2-like protein [Arabidopsis thaliana] >gi 5281023 emb CAB45962.1 (Z97343) GTP-binding RAB2A like protein [Arabidopsis thaliana] Length = 211
479	2023479	Tyr Phospho_Site(465-473)
480	2023480	Tyr Phospho_Site(143-151)
481	2023481	2E-36 >emb CAB39631.1 (AL049481) DNA-directed RNA polymerase [Arabidopsis thaliana] Length = 748
482	2023482	8E-28 >dbj BAA76626.1 (AB019392) muscle specific gene M9 [Homo sapiens] >gi 4689150 gb AAD27784.1 AF077051_1 (AF077051) PTD001 [Homo sapiens] Length = 218
483	2023483	1E-148 >gi 3249095 (AC003114) Contains similarity to dihydrofolate reductase (dfr1) gb L13703 from Schizosaccharomyces pombe. ESTs gb N37567 and gb T43002 come from this gene. [Arabidopsis thaliana] Length = 550
484	2023484	1E-111 >gi 3746809 (AF082882) adenylate kinase [Arabidopsis thaliana] Length = 246
485	2023485	Tyr Phospho_Site(370-378)
486	2023486	7E-61 >gi 549975 (U12858) nucleosome assembly protein I-like protein; similar to mouse nap I, PIR Accession Number JS0707 [Arabidopsis thaliana] Length = 382
487	2023487	1E-105 >sp Q96283 RB1A_ARATH RAS-RELATED PROTEIN RAB11A >gi 2598229 emb CAA70112 (Y08904) Rab11 protein [Arabidopsis thaliana] >gi 5541676 emb CAB51182.1 (AL096859) Rab11 protein [Arabidopsis thaliana] Length = 217
488	2023488	4E-89 >gb AAD25137.1 AC007127_3 (AC007127) ubiquitin protein [Arabidopsis thaliana] Length = 536

489	2023489	Zinc Finger C2h2(1776-1798)
490	2023490	1E-112 >gi 2191174 (AF007270) similar to the peptidase family S16 [Arabidopsis thaliana] Length = 1096
491	2023491	1E-147 >gi 3461837 (AC005315) expansin [Arabidopsis thaliana] >gi 3927842 (AC005727) expansin AtEx6 [Arabidopsis thaliana] Length = 257
492	2023492	1E-173 >gi 3157937 (AC002131) identical to aspartic proteinase cDNA gb U51036 from A. thaliana. ESTs gb N96313, gb T21893, gb R30158, gb T21482, gb T43650, gb R64749, gb R65157, gb T88269, gb T44552, gb T22542, gb T76533, gb T44350, gb Z34591, gb AA728734, gb... Length = 506
493	2023493	4E-43 >dbj BAA25989 (D89051) ERD6 protein [Arabidopsis thaliana] Length = 496
494	2023494	Tyr Phospho Site(419-426)
495	2023495	Tyr Phospho Site(1183-1190)
496	2023496	1E-162 >emb CAA71627 (Y10617) 12-oxophytodienoate reductase [Arabidopsis thaliana] Length = 370
497	2023497	Tyr Phospho Site(1175-1181)
498	2023498	Pkc Phospho Site(18-20)
499	2023499	1E-12 >gi 3834382 (AF033109) syntaxin 8 [Rattus norvegicus] Length = 236
500	2023500	1E-132 >gi 2317729 (AF013627) reversibly glycosylated polypeptide-1 [Arabidopsis thaliana] Length = 357
501	2023501	9E-93 >sp P34091 RL6_MESCR 60S RIBOSOMAL PROTEIN L6 (YL16-LIKE) >gi 280374 pir S28586 ribosomal protein ML16 - common ice plant >gi 19539 emb CAA49175 (X69378) ribosomal protein YL16 [Mesembryanthemum crystallinum] Length =
502	2023502	Pkc Phospho Site(26-28)
503	2023503	3E-11 >gi 4100433 (AF000378) beta-glucosidase [Glycine max] Length = 206
504	2023504	Tyr Phospho Site(1044-1050)
505	2023505	Tyr Phospho Site(659-666)
506	2023506	4E-66 >gi 2443890 (AC002294) similar to NAM (gp X92205 1321924) and CUC2 (gp AB002560 1944132) proteins [Arabidopsis thaliana] Length = 300
507	2023507	8E-24 >gi 3608412 (AF079355) protein phosphatase-2c [Mesembryanthemum crystallinum] Length = 309
508	2023508	Tyr Phospho Site(392-398)
509	2023509	Tyr Phospho Site(184-191)
510	2023510	Tyr Phospho Site(877-883)
511	2023511	8E-22 >gi 2622711 (AE000918) ferripyochelin binding protein [Methanobacterium thermoautotrophicum] Length = 151
512	2023512	Pkc Phospho Site(11-13)
513	2023513	2E-20 >ref NP_005998.1 PZNF216 zinc finger protein 216 >gi 3643809 (AF062346) zinc finger protein 216 splice variant 1 [Homo sapiens] >gi 3643811 (AF062347) zinc finger protein 216 splice variant 2 [Homo sapiens] >gi 3668066 gb AAC61801.1 (AF062072) zinc finger protein 216 [Homo sapiens] Length = 213
514	2023514	Pkc Phospho Site(29-31)
515	2023515	1E-103 >sp Q38912 RAC3_ARATH RAC-LIKE GTP BINDING PROTEIN ARAC3 >gi 1304413 (U43501) Rac-like protein [Arabidopsis thaliana] >gi 2645643 (AF031427) Rho-like GTP binding protein [Arabidopsis thaliana] >gi 2924513 emb CAA17767.1 (AL022023) Rho1Ps homolog/ Rac-like protein [Arabido
516	2023516	4E-46 >emb CAA72716 (Y11987) FPF1 protein [Sinapis alba] Length = 110
517	2023517	1E-119 >emb CAB45987.1 (AL080318) stress-induced protein sti1-like protein [Arabidopsis thaliana] Length = 558

518	2023518	1E-145 >gi 3980379 (AC004561) cyclin, PCNA [Arabidopsis thaliana] Length = 264
519	2023519	1E-66 >emb CAB16514.1 (Z99281) similar to ADP-ribosylation factor; cDNA EST EMBL:C08179 comes from this gene; cDNA EST EMBL:C08337 comes from this gene; cDNA EST EMBL:C09829 comes from this gene; cDNA EST yk291b4.5 comes
520	2023520	Pkc_Phospho_Site(26-28)
521	2023521	2E-45 >emb CAA74401.1 (Y14072) HMG protein [Arabidopsis thaliana] Length = 144
522	2023522	4E-40 >pir S62699 photoassimilate-responsive protein PAR-1b precursor - common tobacco >gi 871487 emb CAA58731 (X83851) mRNA inducible by sucrose and salicylic acid expressed in sugar-accumulating tobacco plants [Ni
523	2023523	Pkc_Phospho_Site(165-167)
524	2023524	2E-60 >gi 3600061 (AF080120) contains similarity to DNA binding proteins [Arabidopsis thaliana] >gi 4850286 emb CAB43042.1 (AL049876) protein [Arabidopsis thaliana] Length = 313
525	2023525	7E-42 >gi 3789911 (AF081802) developmental protein DG1118 [Dictyostelium discoideum] Length = 192
526	2023526	Tyr_Phospho_Site(2-8)
527	2023527	Tyr_Phospho_Site(248-254)
528	2023528	Pkc_Phospho_Site(85-87)
529	2023529	1E-125 >sp P28188 ARA5_ARATH RAS-RELATED PROTEIN ARA-5 >gi 2317906 (U89959) ARA-5 [Arabidopsis thaliana] Length = 258
530	2023530	Zinc_Protease(1367-1376)
531	2023531	1E-127 >gb AAD30573.1 AC007260_4 (AC007260) 50S Ribosomal protein L13 [Arabidopsis thaliana] Length = 241
532	2023532	Pkc_Phospho_Site(53-55)
533	2023533	4E-57 >sp O23760 COMT_CLABR CAFFEIC ACID 3-O-METHYLTRANSFERASE (S-ADENOSYL-L-METHIONINE:CAFFEIC ACID 3-O-METHYLTRANSFERASE) (COMT) >gi 2240207 (AF006009) caffeic acid O-methyltransferase [Clarkia breweri] Length = 370
534	2023534	Tyr_Phospho_Site(884-892)
535	2023535	Pkc_Phospho_Site(55-57)
536	2023536	6E-16 >gi 2281649 (AF003105) AP2 domain containing protein RAP2.12 [Arabidopsis thaliana] Length = 317
537	2023537	6E-34 >emb CAB39533.1 (AJ223758) 54 kDa vacuolar H(+)-ATPase subunit [Sus scrofa] Length = 483
538	2023538	3E-19 >ref NP_005998.1 PZNF216 zinc finger protein 216 >gi 3643809 (AF062346) zinc finger protein 216 splice variant 1 [Homo sapiens] >gi 3643811 (AF062347) zinc finger protein 216 splice variant 2 [Homo sapiens] >gi 3668066 gb AAC61801.1 (AF062072) zinc finger protein 216 [Homo sapiens] Length = 213
539	2023539	Zinc_Finger_C3hc4(1254-1263)
540	2023540	8E-43 >emb CAB40041.1 (AL049524) alpha NAC [Arabidopsis thaliana] Length = 212
541	2023541	3E-64 >emb CAB53477.1 (AJ245900) CAA30374.1 protein [Oryza sativa] Length = 603
542	2023542	1E-93 >pir S42651 hypothetical protein - rape >gi 6065752 emb CAB58175.1 (X74225) pod-specific dehydrogenase SAC25 [Brassica napus] Length = 320
543	2023543	1E-139 >gb AAD25850.1 AC007197_3 (AC007197) cytochrome p450 [Arabidopsis thaliana] Length = 518
544	2023544	1E-124 >emb CAA65988 (X97323) outward rectifying potassium channel KCO1 [Arabidopsis thaliana] >gi 2230761 emb CAA69158 (Y07825) kco1 [Arabidopsis thaliana] Length = 363

545	2023545	Tyr Phospho Site(258-265)
546	2023546	9E-38 >emb CAA74000 (Y13649) homologous to GATA-binding transcription factors [Arabidopsis thaliana] >gi 4895246 gb AAD32831.1 AC007659_13 (AC007659) GATA-binding transcription factor [Arabidopsis thaliana] Le
547	2023547	1E-124 >gb AAD02810 (AF062396) protein phosphatase 2A regulatory subunit isoform B' delta [Arabidopsis thaliana] Length = 477
548	2023548	Tyr Phospho Site(4-11)
549	2023549	1E-32 >dbj BAA22813 (D26015) CND41, chloroplast nucleoid DNA binding protein [Nicotiana tabacum] Length = 502
550	2023550	1E-105 >gi 3860277 (AC005824) ribosomal protein L10 [Arabidopsis thaliana] >gi 4314394 gb AAD15604 (AC006232) ribosomal protein L10A [Arabidopsis thaliana] Length = 222
551	2023551	5E-42 >gb AAD43442.1 AF107837_1 (AF107837) 26S proteasome subunit p40.5 [Homo sapiens] Length = 376
552	2023552	1E-68 >emb CAB36757.1 (AL035523) acid phosphatase-like protein [Arabidopsis thaliana] Length = 260
553	2023553	Pkc Phospho Site(21-23)
554	2023554	0) >gi 3482924 (AC003970) Highly similar to cinnamyl alcohol dehydrogenase, gi 1143445 [Arabidopsis thaliana] Length = 322
555	2023555	4E-94 >gb AAD50055.1 AC007980_20 (AC007980) ATP-dependent metalloprotease [Arabidopsis thaliana] Length = 716
556	2023556	Tyr Phospho Site(1518-1526)
557	2023557	Tyr Phospho Site(254-262)
558	2023558	2E-25 >sp P35559 IDE_RAT INSULIN-DEGRADING ENZYME (INSULYSIN) (INSULINASE) (INSULIN PROTEASE) >gi 347022 pir S29509 insulinase (EC 3.4.99.45) - rat >gi 56492 emb CAA47689 (X67269) insulin-degrading enzyme [Rattus norvegicus]
559	2023559	1E-44 >emb CAA74400.1 (Y14071) HMG protein [Arabidopsis thaliana] >gi 3068715 (AF049236) unknown [Arabidopsis thaliana] Length = 178
560	2023560	1E-109 >gi 2281647 (AF003104) AP2 domain containing protein RAP2.11 [Arabidopsis thaliana] Length = 255
561	2023561	Tyr Phospho Site(300-308)
562	2023562	Pkc Phospho Site(62-64)
563	2023563	9E-61 >emb CAA71502 (Y10477) chloroplast thylakoidal processing peptidase [Arabidopsis thaliana] Length = 340
564	2023564	Tyr Phospho Site(685-692)
565	2023565	1E-12 >gi 3287691 (AC003979) Contains similarity to RING zinc finger protein gb X95455 from Gallus gallus. [Arabidopsis thaliana] Length = 398
566	2023566	Rgd(902-904)
567	2023567	Rgd(1696-1698)
568	2023568	4E-41 >gi 2462833 (AF000657) highly similar to froha and frohb, potential frohc, tumor related protein [Arabidopsis thaliana] Length = 693
569	2023569	Pkc Phospho Site(8-10)
570	2023570	Tyr Phospho Site(1252-1259)
571	2023571	3E-22 >gi 4091808 (AF053307) deacetylindoline 4-O-acetyltransferase [Catharanthus roseus] Length = 439
572	2023572	1E-142 >sp P48422 C861_ARATH CYTOCHROME P450 86A1 (CYPLXXXVI) >gi 940446 emb CAA62082 (X90458) cytochrome p450 [Arabidopsis thaliana] Length = 513
573	2023573	1E-130) >gb AAD50014.1 AC007651_9 (AC007651) glutathione transferase [Arabidopsis thaliana] Length = 220
574	2023574	4E-24 >gb AAD33602.1 AF133302_1 (AF133302) type 2 peroxiredoxin [Brassica rapa subsp. pekinensis] Length = 162
575	2023575	1E-108 >gi 3860277 (AC005824) ribosomal protein L10 [Arabidopsis

		thaliana] >gi 4314394 gb AAD15604 (AC006232) ribosomal protein L10A [Arabidopsis thaliana] Length = 222
576	2023576	Tyr Phospho Site(301-308)
577	2023577	8E-75 >emb CAA17547.1 (AL021960) photosystem II oxygen-evolving complex protein 3-like [Arabidopsis thaliana] >gi 3402748 emb CAA20194.1 (AL031187) photosystem II oxygen-evolving complex protein 3-like [Arabidopsis thaliana] Length = 223
578	2023578	Tyr Phospho Site(49-56)
579	2023579	1E-83 >emb CAA18743.1 (AL022604) NAD+ dependent isocitrate dehydrogenase subunit 1 [Arabidopsis thaliana] Length = 367
580	2023580	Pkc Phospho Site(2-4)
581	2023581	5E-40 >pir S52995 arabinogalactan-like protein - loblolly pine >gi 607774 (U09556) arabinogalactan-like protein [Pinus taeda] Length = 264
582	2023582	4E-23 >emb CAA10616 (AJ132240) eukaryotic translation initiation factor 5 [Zea mays] Length = 451
583	2023583	2E-65 >sp P29545 EF1D_ORYSA ELONGATION FACTOR 1-BETA' (EF-1-BETA') >gi 322851 pir S29224 translation elongation factor eEF-1 beta' chain - rice >gi 218161 dbj BAA02253 (D12821) elongation factor 1 beta' [Oryza sativa] Length = 223
584	2023584	1E-36 >gb AAF00645.1 AC009540_22 (AC009540) cationic amino acid transporter [Arabidopsis thaliana] Length = 614
585	2023585	1E-123 >gi 3152563 (AC002986) Similar to myb-related transcription factors e.g., gb X98308. EST gb T22093 and gb T22697 come from this gene. [Arabidopsis thaliana] Length = 327
586	2023586	9E-13 >emb CAB10221.1 (Z97336) elicitor like protein [Arabidopsis thaliana] Length = 158
587	2023587	1E-100 >gb AAD35009.1 AF144391_1 (AF144391) thioredoxin-like 5 [Arabidopsis thaliana] Length = 185
588	2023588	Rgd(1535-1537)
589	2023589	1E-105 >gi 2262173 (AC002329) NADPH thioredoxin reductase [Arabidopsis thaliana] Length = 383
590	2023590	Tyr Phospho Site(1491-1497)
591	2023591	Tyr Phospho Site(966-972)
592	2023592	2E-56 >sp Q06138 MO25_MOUSE MO25 PROTEIN >gi 2143483 pir 57997 hypothetical calcium-binding protein - mouse >gi 262934 bbs 121784 (S51858) Ca2+ binding protein [mice, embryos, Peptide, 341 aa] [Mus sp.] Length = 341
593	2023593	4E-99 >gi 3822225 (AF079183) RING-H2 finger protein RHG1a [Arabidopsis thaliana] Length = 190
594	2023594	5E-98 >dbj BAA31144 (AB010916) response reactor2 [Arabidopsis thaliana] >gi 4678318 emb CAB41129.1 (AL049658) response reactor2 [Arabidopsis thaliana] Length = 184
595	2023595	1E-122 >gi 1046225 (U21952) ethylene response sensor [Arabidopsis thaliana] >gi 2623308 (AC002409) ethylene response sensor (ERS) [Arabidopsis thaliana] >gi 1584365 prf 2122405A ERS gene [Arabidopsis thaliana] Length = 613
596	2023596	3E-28 >gi 2494114 (AC002376) Contains similarity to Daucus glycine-rich cell wall protein (gb D29974). EST gb R29840 comes from this gene. [Arabidopsis thaliana] Length = 212
597	2023597	Tyr Phospho Site(780-786)
598	2023598	2E-80 >emb CAA09198 (AJ010459) RNA helicase [Arabidopsis thaliana] Length = 145
599	2023599	7E-27 >gb AAD46402.1 AF096246_1 (AF096246) ethylene-responsive transcriptional coactivator [Lycopersicon esculentum] Length = 146
600	2023600	Pkc Phospho Site(151-153)
601	2023601	2E-82 >gb AAD27618.1 AF124376_1 (AF124376) 30S ribosomal protein S7

		[Brassica napus] >gi 5881740 dbj BAA84431.1 (AP000423) ribosomal protein S7 [Arabidopsis thaliana] >gi 5881755 dbj BAA84446.1 (AP000423) ribosomal protein S7 [Arabidopsis thaliana] Length = 155
602	2023602	2E-79 >gb AAD14462 (AC005275) glycosylation enzyme [Arabidopsis thaliana] Length = 448
603	2023603	4E-98 >dbj BAA74528 (AB016471) ARR1 protein [Arabidopsis thaliana] Length = 669
604	2023604	5E-74 >gi 3169883 (AF033194) dehydroquinase dehydratase/shikimate:NADP oxidoreductase [Lycopersicon esculentum] >gi 3169888 (AF034411) dehydroquinase dehydratase/shikimate:NADP oxidoreductase [Lycopersicon esculentum] Length = 545
605	2023605	Tyr Phospho Site(382-390)
606	2023606	Tyr Phospho Site(1085-1092)
607	2023607	Tyr Phospho Site(538-545)
608	2023608	2E-69 >gb AAD21706.1 (AC007048) tyrosine transaminase [Arabidopsis thaliana] Length = 462
609	2023609	Tyr Phospho Site(216-223)
610	2023610	Pkc Phospho Site(10-12)
611	2023611	1E-35 >gb AAD45979.1 (AF115334) MenG [Pseudomonas fluorescens] Length = 163
612	2023612	9E-23 >dbj BAA32422 (AB008107) ethylene responsive element binding factor 5 [Arabidopsis thaliana] Length = 300
613	2023613	2E-90 >pir S71219 cytosolic cyclophilin ROC3 - Arabidopsis thaliana >gi 1305455 (U40399) cytosolic cyclophilin [Arabidopsis thaliana] >gi 4581104 gb AAD24594.1 AC005825_1 (AC005825) cytosolic cyclophilin (ROC3) [Arabidopsis thaliana] Length = 173
614	2023614	Tyr Phospho Site(78-86)
615	2023615	Pkc Phospho Site(12-14)
616	2023616	Tyr Phospho Site(772-780)
617	2023617	1E-106 >emb CAB45054.1 (AL078637) HSP90-like protein [Arabidopsis thaliana] Length = 823
618	2023618	1E-101 >gi 4056469 (AC005990) Strong similarity to gb M95166 ADP-ribosylation factor from Arabidopsis thaliana. ESTs gb Z25826, gb R90191, gb N65697, gb AA713150, gb T46332, gb AA040967, gb AA712956, gb T46403, gb T46050, gb AI100391 and gb Z25043 come from t... Length = 188
619	2023619	Tyr Phospho Site(9-16)
620	2023620	3E-44 >gi 3201632 (AC004669) 2A6 protein [Arabidopsis thaliana] Length = 358
621	2023621	1E-113 >emb CAB10222.1 (Z97336) carnitine racemase like protein [Arabidopsis thaliana] Length = 240
622	2023622	1E-63 >gi 3341698 (AC003672) blue copper-binding protein II [Arabidopsis thaliana] Length = 202
623	2023623	1E-108 >sp Q96558 UGDH_SOYBN UDP-GLUCOSE 6-DEHYDROGENASE (UDP-GLC DEHYDROGENASE) (UDP-GLCDH) (UDPGDH) >gi 1518540 (U53418) UDP-glucose dehydrogenase [Glycine max] Length = 480
624	2023624	Tyr Phospho Site(515-522)
625	2023625	Tyr Phospho Site(1716-1723)
626	2023626	2E-16 >emb CAA84724.1 (Z35663) similar to ribonucleoprotein; cDNA EST yk222a11.3 comes from this gene; cDNA EST yk222a11.5 comes from this gene; cDNA EST yk432f10.3 comes from this gene; cDNA EST yk432f10.5 comes from this gene; cDNA EST yk497a8.3 ... Length = 307
627	2023627	2E-57 >gi 3482933 (AC003970) Similar to cdc2 protein kinases [Arabidopsis thaliana] Length = 967
628	2023628	Tyr Phospho Site(4-12)
629	2023629	4E-92 >gi 3201969 (AF068332) submergence induced protein 2A [Oryza

		sativa] Length = 198
630	2023630	1E-110 >gb AAD41977.1 AC006438_9 (AC006438) unknown protein [Arabidopsis thaliana] Length = 203
631	2023631	Tyr_Phospho_Site(983-990)
632	2023632	1E-106 >gi 3482931 (AC003970) germin-like protein [Arabidopsis thaliana] Length = 219
633	2023633	4E-68 >gi 4193388 (AF091455) translationally controlled tumor protein [Hevea brasiliensis] Length = 168
634	2023634	5E-23 >gi 3193325 (AF069299) contains similarity to pectinesterases [Arabidopsis thaliana] Length = 209
635	2023635	2E-45 >emb CAB52425.1 (AL109770) similar to yeast vacuolar sorting protein VPS29/PEP11 [Schizosaccharomyces pombe] Length = 187
636	2023636	9E-16 >sp P53173 ERV4_YEAST ER-DERIVED VESICLES PROTEIN ERV14 >gi 2132531 pir S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae) >gi 1322550 emb CAA96756 (Z72576) ORF YGL054c [Saccharomyces cerevisiae] Length = 138
637	2023637	1E-126 >gi 3415113 (AF081201) villin 1 [Arabidopsis thaliana] Length = 910
638	2023638	1E-125 >pir S58282 dTDP-glucose 4-6-dehydratases homolog - Arabidopsis thaliana >gi 928932 emb CAA89205 (Z49239) homolog of dTDP-glucose 4-6-dehydratases [Arabidopsis thaliana] >gi 1585435 prf 2124427B diamide resistance gene [Arabidopsis thaliana] Length = 445
639	2023639	Tyr_Phospho_Site(1102-1110)
640	2023640	2E-30 >sp Q01264 HYUC_PSESN HYDANTOIN UTILIZATION PROTEIN C (ORF4) >gi 151284 (M72717) DL-hydantoinase [Pseudomonas sp.] >gi 216833 dbj BAA01379 (D10494) N-carbamyl-L-amino acid amidohydrolase [Pseudomonas sp.] Length = 414
641	2023641	Tyr_Phospho_Site(127-134)
642	2023642	Tyr_Phospho_Site(407-413)
643	2023643	1E-155 >gb AAD21710.1 (AC007048) protein phosphatase 2C [Arabidopsis thaliana] Length = 290
644	2023644	4E-97 >gi 862640 (U20182) MADS-box protein AGL11 [Arabidopsis thaliana] >gi 4538999 emb CAB39620.1 (AL049481) MADS-box protein AGL11 [Arabidopsis thaliana] Length = 230
645	2023645	1E-127 >gi 3894171 (AC005312) glutathione s-transferase [Arabidopsis thaliana] Length = 221
646	2023646	1E-120 >sp Q39222 RB1B_ARATH RAS-RELATED PROTEIN RAB11 >gi 2118459 pir S59942 small GTP-binding protein Rab11 - Arabidopsis thaliana >gi 451860 (L18883) small GTP-binding protein [Arabidopsis thaliana] Length = 216
647	2023647	Tyr_Phospho_Site(162-168)
648	2023648	7E-29 >dbj BAA22813 (D26015) CND41, chloroplast nucleoid DNA binding protein [Nicotiana tabacum] Length = 502
649	2023649	1E-34 >dbj BAA12797 (D85381) cytochrome c oxidase subunit Vb precursor [Oryza sativa] Length = 169
650	2023650	Pkc_Phospho_Site(60-62)
651	2023651	Tyr_Phospho_Site(927-934)
652	2023652	1E-128 >gb AAD20681 (AC006283) similar to protein Htf9C [Arabidopsis thaliana] Length = 850
653	2023653	1E-117 >gb AAD22643.1 AC007138_7 (AC007138) protein transport factor [Arabidopsis thaliana] Length = 856
654	2023654	Tyr_Phospho_Site(951-957)
655	2023655	Pkc_Phospho_Site(31-33)
656	2023656	8E-23 >emb CAB50433.1 (AJ248287) hypothetical DEHYDROGENASE [Pyrococcus abyssi] Length = 333

657	2023657	1E-129 >sp Q08770 RL10_ARATH 60S RIBOSOMAL PROTEIN L10 (WILM'S TUMOR SUPPRESSOR PROTEIN HOMOLOG) >gi 478401 pir JQ2244 ribosomal protein L10.e, cytosolic - Arabidopsis thaliana >gi 17682 emb CAA78856 (Z15157) Wilm's tumor suppressor homologue [Arabidopsis thaliana] Length = 220
658	2023658	6E-22 >gb AAD32844.1 AC007658_3 (AC007658) thioredoxin-like protein [Arabidopsis thaliana] Length = 130
659	2023659	1E-141 >emb CAB41166.1 (AL049659) cytochrome P450-like protein [Arabidopsis thaliana] Length = 490
660	2023660	Pkc Phospho Site(177-179)
661	2023661	7E-92 >gi 4056504 (AC005896) zinc finger protein [Arabidopsis thaliana] Length = 178
662	2023662	Tyr Phospho Site(441-448)
663	2023663	Tyr Phospho Site(1407-1415)
664	2023664	2E-60 >gi 1532175 (U63815) similar to protein disulfide isomerase [Arabidopsis thaliana] Length = 132
665	2023665	1E-128 >emb CAB10215.1 (Z97336) ankyrin like protein [Arabidopsis thaliana] Length = 936
666	2023666	Tyr Phospho Site(764-772)
667	2023667	1E-107 >emb CAB52747.1 (AJ245629) photosystem I subunit III precursor [Arabidopsis thaliana] Length = 221
668	2023668	Tyr Phospho Site(146-152)
669	2023669	1E-112 >gi 3065835 (AF058800) methyltransferase [Arabidopsis thaliana] Length = 504
670	2023670	Tyr Phospho Site(910-918)
671	2023671	Tyr Phospho Site(1058-1064)
672	2023672	Tyr Phospho Site(377-383)
673	2023673	2E-33 >gi 4097549 (U64907) ATPF4 [Arabidopsis thaliana] Length = 179
674	2023674	1E-119 >sp P41916 RAN1_ARATH GTP-BINDING NUCLEAR PROTEIN RAN-1 >gi 495729 (L16789) small ras-related protein [Arabidopsis thaliana] >gi 2058278 emb CAA66047 (X97379) atran1 [Arabidopsis thaliana] Length = 221
675	2023675	1E-105 >sp P22953 HS71_ARATH HEAT SHOCK COGNATE 70 KD PROTEIN 1 >gi 1072473 pir S46302 heat shock cognate protein 70-1 - Arabidopsis thaliana >gi 397482 emb CAA52684 (X74604) heat shock protein 70 cognate [Arabidopsis thaliana] Length = 651
676	2023676	2E-89 >gb AAD39282.1 AC007576_5 (AC007576) Similar to DNA-binding proteins [Arabidopsis thaliana] Length = 487
677	2023677	1E-127 >gi 4056505 (AC005896) nodulin-like protein [Arabidopsis thaliana] Length = 357
678	2023678	1E-135 >gi 886116 (U27609) TCH4 protein [Arabidopsis thaliana] >gi 2952473 (AF051338) xyloglucan endotransglycosylase related protein [Arabidopsis thaliana] Length = 284
679	2023679	2E-90 >sp O23255 SAHH_ARATH ADENOSYLHOMOCYSTEINASE (S-ADENOSYL-L-HOMOCYSTEINE HYDROLASE) (ADOHCYASE) >gi 2244750 emb CAB10173.1 (Z97335) adenosylhomocysteinase [Arabidopsis thaliana] >gi 3088579 gb AAC14714.1 (AF059581) S-adenosyl-L-homocysteine hydrolase [Arabidopsis thaliana] Length = 485
680	2023680	9E-23 >dbj BAA32422 (AB008107) ethylene responsive element binding factor 5 [Arabidopsis thaliana] Length = 300
681	2023681	Tyr Phospho Site(304-312)
682	2023682	Tyr Phospho Site(654-660)
683	2023683	2E-58 >sp Q43434 VATL_GOSHI VACUOLAR ATP SYNTHASE 16 KD PROTEOLIPID SUBUNIT >gi 755148 (U13669) vacuolar H+-ATPase proteolipid (16 kDa) subunit [Gossypium hirsutum] >gi 4519415 dbj BAA75542.1 (AB024275)

		vacuolar H ⁺ -ATPase c subunit [Citrus unshiu] Length = 165
684	2023684	1E-106 >pir S50767 protein kinase - rice >gi 450300 (L27821) protein kinase [Oryza sativa] Length = 824
685	2023685	6E-14 >sp Q28891 S5A1_MACFA 3-OXO-5-ALPHA-STEROID 4-DEHYDROGENASE 1 (STEROID 5-ALPHA-REDUCTASE 1) (SR TYPE 1) >gi 999036 bbs 164548 (S77162) steroid 5 alpha-reductase type 1 isoenzyme, SR type 1 [Cynomolgus monkeys, prostate, Peptide, 263 aa] [Macaca fascicularis] Length = 263
686	2023686	1E-131 >gb AAC34217.1 (AC004411) alcohol dehydrogenase [Arabidopsis thaliana] Length = 257
687	2023687	Tyr Phospho Site(146-152)
688	2023688	2E-72 >emb CAB44322.1 (AL078606) phospholipase D-gamma [Arabidopsis thaliana] Length = 866
689	2023689	8E-97 >emb CAB53034.1 (AJ245867) photosystem I subunit XI precursor [Arabidopsis thaliana] Length = 219
690	2023690	1E-133 >sp O80585 MTHR_ARATH PROBABLE METHYLENETETRAHYDROFOLATE REDUCTASE >gi 3212869 (AC004005) unknown protein [Arabidopsis thaliana] Length = 606
691	2023691	Tyr Phospho Site(501-508)
692	2023692	6E-26 >gb AAD40017.1 AF150111_1 (AF150111) small zinc finger-like protein [Arabidopsis thaliana] Length = 93
693	2023693	1E-101 >gi 4056469 (AC005990) Strong similarity to gb M95166 ADP-ribosylation factor from Arabidopsis thaliana. ESTs gb Z25826, gb R90191, gb N65697, gb AA713150, gb T46332, gb AA040967, gb AA712956, gb T46403, gb T46050, gb A1100391 and gb Z25043 come from t... Length = 188
694	2023694	Zinc Protease(160-169)
695	2023695	3E-94 >emb CAB36847.1 (AL035528) DnaJ-like protein [Arabidopsis thaliana] Length = 197
696	2023696	Tyr Phospho Site(1062-1069)
697	2023697	1E-83 >sp P35132 UBC9_ARATH UBIQUITIN-CONJUGATING ENZYME E2-17 KD 9 (UBIQUITIN-PROTEIN LIGASE 9) (UBIQUITIN CARRIER PROTEIN 9) (UBCAT4B) >gi 421857 pir S32674 ubiquitin—protein ligase (EC 6.3.2.19) UBC9 - Arabidopsis thaliana >gi 297884 emb CAA78714 (Z14990) ubiquitin conjugating enzyme homolog [Arabidopsis thaliana] >gi 349211 (L00639) ubiquitin conjugating enzyme [Arabidopsis thaliana] >gi 600391 emb CAA51201 (X72626) ubiquitin conjugating enzyme E2 [Arabidopsis thaliana] >gi 4455355 emb CAB36765.1 (AL035524) ubiquitin-protein ligase UBC9 [Arabidopsis thaliana] Length = 148
698	2023698	2E-47 >emb CAA09200 (AJ010461) RNA helicase [Arabidopsis thaliana] Length = 363
699	2023699	Tyr Phospho Site(1315-1322)
700	2023700	3E-86 >gb AAD22122.1 AC006224_4 (AC006224) isopropylmalate dehydratase [Arabidopsis thaliana] Length = 256
701	2023701	9E-11 >pir S59397 probable membrane protein YLR251w - yeast (Saccharomyces cerevisiae) >gi 662333 (U20865) Ylr251wp [Saccharomyces cerevisiae] Length = 197
702	2023702	1E-113 >sp O23755 EF2_BETVU ELONGATION FACTOR 2 (EF-2) >gi 2369714 emb CAB09900 (Z97178) elongation factor 2 [Beta vulgaris] Length = 843
703	2023703	8E-46 >pir A39634 probable cell cycle control protein crn - fruit fly (Drosophila melanogaster) >gi 2827496 emb CAA15705.1 (AL009195) EG:30B8.1 [Drosophila melanogaster] Length = 702
704	2023704	Tyr Phospho Site(1307-1314)
705	2023705	1E-145 >gb AAD46682.1 AF170910_1 (AF170910) SYNC2 protein [Arabidopsis thaliana] Length = 638
706	2023706	1E-65 >gi 3341698 (AC003672) blue copper-binding protein II

		[Arabidopsis thaliana] Length = 202
707	2023707	Rgd(993-995)
708	2023708	Tyr_Phospho_Site(94-101)
709	2023709	Tyr_Phospho_Site(1050-1057)
710	2023710	1E-107 >gb AAD39612.1 AC007454_11 (AC007454) Similar to gb X92204 NAM gene product from Petunia hybrida. ESTs gb H36656 and gb AA651216 come from this gene. [Arabidopsis thaliana] Length = 557
711	2023711	7E-88 >gb AAD27909.1 AC007213_7 (AC007213) receptor protein kinase [Arabidopsis thaliana] Length = 851
712	2023712	2E-89 >dbj BAA18577 (D90915) peptide chain release factor [Synechocystis sp.] Length = 288
713	2023713	4E-54 >gb AAD21451.1 (AC007017) DNA-binding protein [Arabidopsis thaliana] Length = 145
714	2023714	Tyr_Phospho_Site(7-14)
715	2023715	Tyr_Phospho_Site(467-473)
716	2023716	Tyr_Phospho_Site(185-191)
717	2023717	6E-48 >gb AAD39312.1 AC007258_1 (AC007258) Similar to glutathione transferase [Arabidopsis thaliana] Length = 234
718	2023718	8E-17 >sp Q42534 PME2_ARATH PECTINESTERASE 2 (PECTIN METHYLESTERASE 2) (PE 2) >gi 2129667 pir PC4168 pectinesterase (EC 3.1.1.11) 2 precursor - Arabidopsis thaliana (fragment) >gi 903894 (U25649) ATPME2 precursor [Arabidopsis thaliana] Length = 582
719	2023719	Tyr_Phospho_Site(1205-1211)
720	2023720	Tyr_Phospho_Site(297-304)
721	2023721	1E-103 >sp Q96252 ATP4_ARATH ATP SYNTHASE DELTA' CHAIN, MITOCHONDRIAL PRECURSOR >gi 1655484 dbj BAA13601 (D88376) delta-prime subunit of mitochondrial F1-ATPase [Arabidopsis thaliana] Length = 203
722	2023722	9E-59 >emb CAB39656.1 (AL049483) nitrogen fixation like protein [Arabidopsis thaliana] Length = 224
723	2023723	2E-27 >gi 2984333 (AE000774) Na(+) dependent transporter (Sbf family) [Aquifex aeolicus] Length = 297
724	2023724	Tyr_Phospho_Site(780-786)
725	2023725	2E-45 >gb AAD22286.1 AC006920_10 (AC006920) reverse transcriptase [Arabidopsis thaliana] Length = 1311
726	2023726	4E-44 >emb CAA63223 (X92491) TOM20 [Solanum tuberosum] Length = 204
727	2023727	1E-23 >emb CAB10456.1 (Z97342) nuclear antigen homolog [Arabidopsis thaliana] Length = 355
728	2023728	1E-82 >dbj BAA06384 (D30719) ERD15 protein [Arabidopsis thaliana] >gi 3241941 (AC004625) dehydration-induced protein ERD15 [Arabidopsis thaliana] >gi 3894181 (AC005662) ERD15 protein [Arabidopsis thaliana] Length = 163
729	2023729	6E-24 >gb AAD24601.1 AC005825_8 (AC005825) reverse transcriptase [Arabidopsis thaliana] Length = 1319
730	2023730	1E-36 >emb CAB16764.1 (Z99707) heat shock transcription factor HSF4 [Arabidopsis thaliana] >gi 3256070 emb CAA74398 (Y14069) Heat Shock Factor 4 [Arabidopsis thaliana] Length = 284
731	2023731	1E-68 >gb AAD25624.1 AC005287_26 (AC005287) Similar to phosphoprotein phosphatase 2A regulatory subunit [Arabidopsis thaliana] Length = 535
732	2023732	1E-114 >gb AAD41426.1 AC007727_15 (AC007727) Identical to gb Y13173 Arabidopsis thaliana mRNA for proteasome subunit. EST gb T76747 comes from this gene. Length = 204
733	2023733	1E-105 >sp P41127 RL13_ARATH 60S RIBOSOMAL PROTEIN L13 (BBC1 PROTEIN HOMOLOG) >gi 480787 pir S37271 ribosomal protein L13 - Arabidopsis thaliana >gi 404166 emb CAA53005 (X75162) BBC1 protein

		[Arabidopsis thaliana] Length = 206
734	2023734	Tyr_Phospho_Site(199-205)
735	2023735	4E-41 >emb CAB44393.1 (AL078610) hydrolase [Streptomyces coelicolor] Length = 269
736	2023736	5E-29 >gb AAD56248.1 AF186273_1 (AF186273) leucine-rich repeats containing F-box protein FBL3 [Homo sapiens] Length = 423
737	2023737	Tyr_Phospho_Site(1188-1195)
738	2023738	5E-63 >gi 3834306 (AC005679) EST gb R65024 comes from this gene. [Arabidopsis thaliana] Length = 156
739	2023739	1E-78 >gi 1707018 (U78721) CutA isolog [Arabidopsis thaliana] Length = 182
740	2023740	1E-164 >gb AAD17364 (AF128396) Arabidopsis thaliana flavin-type blue-light photoreceptor (SW:Q43125) (Pfam: PF00875, Score=765.2, E=2.6e-226, N=1) [Arabidopsis thaliana] Length = 702
741	2023741	9E-14 >ref NP_003913.1 PHERC1 guanine nucleotide exchange factor p532 >gi 1477565 (U50078) p532 [Homo sapiens] Length = 4861
742	2023742	1E-133 >emb CAA65053 (X95738) proline transporter 2 [Arabidopsis thaliana] Length = 439
743	2023743	6E-93 >gb AAD39312.1 AC007258_1 (AC007258) Similar to glutathione transferase [Arabidopsis thaliana] Length = 234
744	2023744	Tyr_Phospho_Site(748-755)
745	2023745	1E-120 >gb AAC24832 (AF061518) manganese superoxide dismutase [Arabidopsis thaliana] Length = 231
746	2023746	3E-83 >emb CAB45986.1 (AL080318) protein [Arabidopsis thaliana] Length = 206
747	2023747	3E-22 >gi 895613 (L43505) CASP gene product [Gallus gallus] Length = 675
748	2023748	4E-39 >gb AAD21699.1 (AC004793) Contains reverse transcriptase domain (rvt) PF 00078. [Arabidopsis thaliana] Length = 1253
749	2023749	1E-124 >emb CAA19720.1 (AL030978) GH3 like protein [Arabidopsis thaliana] Length = 612
750	2023750	1E-69 >emb CAB36546.1 (AL035440) DNA binding protein [Arabidopsis thaliana] Length = 427
751	2023751	3E-75) >gi 1707022 (U78721) proline-rich protein isolog [Arabidopsis thaliana] Length = 239
752	2023752	1E-122 >gb AAD17428 (AC006284) methyltransferase [Arabidopsis thaliana] Length = 619
753	2023753	3E-15 >gi 2252854 (AF013294) similar to auxin-induced protein [Arabidopsis thaliana] Length = 122
754	2023754	1E-101 >gi 2444176 (U94782) unconventional myosin [Helianthus annuus] Length = 1260
755	2023755	Tyr_Phospho_Site(661-668)
756	2023756	7E-97 >gb AAD15400 (AC006223) integral membrane protein [Arabidopsis thaliana] Length = 429
757	2023757	1E-120 >sp P42761 GTH3_ARATH GLUTATHIONE S-TRANSFERASE ERD13 (GST CLASS PHI) >gi 481822 pir S39542 probable glutathione transferase (EC 2.5.1.18) (clone ERD13) - Arabidopsis thaliana >gi 497789 dbj BAA04554 (D17673) glutathio
758	2023758	1E-114) >gi 1707015 (U78721) protein phosphatase 2C isolog [Arabidopsis thaliana] Length = 380
759	2023759	1E-108 >gb AAD24598.1 AC005825_5 (AC005825) chloroplast outer membrane protein 86,also very similar to GTP-inding protein from pea (GB:L36857) [Arabidopsis thaliana] Length = 1206
760	2023760	1E-82 >emb CAA16964 (AL021811) H+-transporting ATP synthase chain9 - like protein [Arabidopsis thaliana] >gi 5730141 emb CAB52473.1

		(AJ245574) ATP synthase beta chain precursor (subunit II) [Arabidopsis thaliana] Length = 219
761	2023761	3E-47 >emb CAA68848 (Y07563) hin1 [Nicotiana tabacum] Length = 221
762	2023762	9E-51 >sp P28342 GTT1_DIACA GLUTATHIONE S-TRANSFERASE 1 (SR8) (GST CLASS-THETA) >gi 99589 pir S16604 glutathione transferase (EC 2.5.1.18) CARSR8 - clove pink >gi 18330 emb CAA41279 (X58390) glutathione s-transferase [Dianthus caryophyllus] >gi 167968 (M64268) glutathione transferase [Dianthus caryophyllus] Length = 221
763	2023763	Tyr Phospho Site(192-199)
764	2023764	Tyr Phospho Site(1388-1396)
765	2023765	1E-38 >emb CAB40579.1 (AJ133639) SAH7 protein [Arabidopsis thaliana] Length = 159
766	2023766	4E-17 >ref NP_003554.1 PSPOP speckle-type POZ protein >gi 2695708 emb CAA04199 (AJ000644) SPOP [Homo sapiens] Length = 374
767	2023767	Pkc Phospho Site(22-24)
768	2023768	3E-31 >sp P81650 BGAL_PSBAT BETA-GALACTOSIDASE (LACTASE) >gi 4079639 emb CAA10470 (AJ131635) beta-galactosidase [psychrophilic bacterium TAE 79] Length = 1039
769	2023769	1E-123 >gi 871782 (L43081) pEARL1 4 gene product [Arabidopsis thaliana] Length = 766
770	2023770	2E-77 >gi 3386612 (AC004665) DNA-binding protein, dbp [Arabidopsis thaliana] Length = 190
771	2023771	1E-29 >sp P42763 DH14_ARATH DEHYDRIN ERD14 >gi 556474 dbj BAA04569 (D17715) ERD14 protein [Arabidopsis thaliana] Length = 185
772	2023772	8E-13 >emb CAA88860.1 (Z49068) similar to GTP-binding protein; cDNA EST EMBL:M89111 comes from this gene; cDNA EST EMBL:D27709 comes from this gene; cDNA EST EMBL:D27708 comes from this gene; cDNA EST EMBL:D73788 comes from this gene; cDNA EST yk3... Length = 556
773	2023773	1E-107 >gb AAC34243.1 (AC004411) pto kinase [Arabidopsis thaliana] Length = 365
774	2023774	9E-88 >gi 3075394 (AC004484) beta-ketoacyl-CoA synthase [Arabidopsis thaliana] >gi 3559809 emb CAA09311 (AJ010713) fiddlehead protein [Arabidopsis thaliana] Length = 550
775	2023775	Tyr Phospho Site(428-434)
776	2023776	1E-125 >emb CAB45880.1 (AL080282) protein [Arabidopsis thaliana] Length = 1396
777	2023777	5E-73 >sp P52810 RS9_PODAN 40S RIBOSOMAL PROTEIN S9 (S7) >gi 1321917 emb CAA65433 (X96613) cytoplasmic ribosomal protein S7 [Podospira anserina] Length = 190
778	2023778	1E-138 >gi 1066499 (L37606) NADH-dependent glutamate synthase [Medicago sativa] Length = 2194
779	2023779	4E-37 >gb AAD19788 (AC006528) zinc-finger protein, 5' partial [Arabidopsis thaliana] Length = 626
780	2023780	1E-10 >gi 3600032 (AF080119) contains similarity to tropomyosin (Pfam: Tropomyosin.hmm, score: 14.57) and ATP synthase (Pfam: ATP-synt_B.hmm, score: 10.89) [Arabidopsis thaliana] Length = 466
781	2023781	9E-86 >gi 2924779 (AC002334) 3-ketoacyl-CoA thiolase [Arabidopsis thaliana] >gi 2981616 dbj BAA25248 (AB008854) 3-ketoacyl-CoA thiolase [Arabidopsis thaliana] >gi 2981618 dbj BAA25249 (AB008855) 3-ketoacyl
782	2023782	2E-91 >emb CAB16762.1 (Z99707) caltractin-like protein [Arabidopsis thaliana] Length = 167
783	2023783	3E-50 >gb AAD21025 (AF106939) 1,4-benzoquinone reductase [Phanerochaete chrysosporium] Length = 201
784	2023784	Tyr Phospho Site(1296-1304)

785	2023785	Tyr Phospho Site(290-296)
786	2023786	2E-52 >gb AAD22344.1 AC006592_1 (AC006592) anthocyanidin-3-glucoside rhamnosyltransferase, 3' partial [Arabidopsis thaliana] Length = 414
787	2023787	Tyr Phospho Site(49-56)
788	2023788	1E-70 >emb CAB41005.1 (AL049640) blue copper-binding protein, 15K (lamin) [Arabidopsis thaliana] Length = 141
789	2023789	8E-25 >sp P73689 SPPA_SYNY3 PROTEASE IV HOMOLOG (ENDOPEPTIDASE IV) >gi 1652816 dbj BAA17735 (D90908) protease IV [Synechocystis sp.] Length = 610
790	2023790	1E-120 >sp Q42599 NUIM_ARATH NADH-UBIQUINONE OXIDOREDUCTASE 23 KD SUBUNIT PRECURSOR (COMPLEX I-23KD) (CI-23KD) >gi 1076356 pir S52380 NADH dehydrogenase (EC 1.6.99.3) - Arabidopsis thaliana >gi 666977 emb CAA59061 (X84318) NADH dehydrogenase [Arabidopsis thaliana] >gi 3152573
791	2023791	4E-91 >gb AAD44761.1 AF144752_1 (AF144752) 40S ribosomal protein S7 homolog [Brassica oleracea] Length = 191
792	2023792	1E-121 >pir S36884 ketol-acid reductoisomerase (EC 1.1.1.86) - Arabidopsis thaliana >gi 402552 emb CAA49506 (X69880) ketol-acid reductoisomerase [Arabidopsis thaliana] Length = 591
793	2023793	Pkc Phospho Site(29-31)
794	2023794	8E-53 >gi 4220474 (AC006069) myosin heavy chain [Arabidopsis thaliana] Length = 629
795	2023795	1E-140 >sp O64637 C7C2_ARATH CYTOCHROME P450 76C2 >gi 2979549 (AC003680) 7-ethoxycoumarin O-deethylase [Arabidopsis thaliana] Length = 512
796	2023796	1E-77 >emb CAA96435 (Z71753) pectin methylesterase [Nicotiana plumbaginifolia] Length = 315
797	2023797	4E-79 >emb CAB41928.1 (AL049751) short-chain alcohol dehydrogenase like protein [Arabidopsis thaliana] Length = 263
798	2023798	3E-27 >ref NP_006818.1 PTMP21 transmembrane trafficking protein >gi 3915893 sp P49755 TM21_HUMAN TRANSMEMBRANE PROTEIN TMP21 PRECURSOR (S3111125) (S311125) >gi 1359886 emb CAA66071 (X97442) transmembrane protein [Homo sapiens] >gi 1407826 (U61734) protein trafficking protein [Homo sapiens] >gi 3288463 emb CAA06213.1 (AJ004913) integral membrane protein, Tmp21-l (p23) [Homo sapiens] >gi 4885697 gb AAD31941.1 AC007055_6 (AC007055) TMP21 [Homo sapiens] Length = 219
799	2023799	Tyr Phospho Site(250-257)
800	2023800	8E-19 >gi 3193325 (AF069299) contains similarity to pectinesterases [Arabidopsis thaliana] Length = 209
801	2023801	Tyr Phospho Site(236-242)
802	2023802	1E-147 >emb CAB41122.1 (AL049657) proteasome regulatory subunit [Arabidopsis thaliana] Length = 406
803	2023803	2E-49 >emb CAB00039.1 (Z75712) Similarity to S. Pombe BEM1/BUD5 suppressor; cDNA EST EMBL:Z14470 comes from this gene; cDNA EST yk482d4.3 comes from this gene; cDNA EST yk482d4.5 comes from this gene [Caenorhabditis elegans] Length = 405
804	2023804	3E-77 >emb CAB38828.1 (AL035679) proton pump [Arabidopsis thaliana] Length = 843
805	2023805	Pkc Phospho Site(74-76)
806	2023806	Pkc Phospho Site(147-149)
807	2023807	2E-97 >sp P49177 GBB_ARATH GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT >gi 557694 (U12232) GTP binding protein beta subunit [Arabidopsis thaliana] >gi 3096915 emb CAA18825.1 (AL023094) GTP binding protein beta subunit [A
808	2023808	2E-79 >dbj BAA13947 (D89341) luminal binding protein [Arabidopsis

		thaliana] Length = 669
809	2023809	5E-79 >emb CAA73063.1 (Y12459) cytosolic glutamine synthetase [Brassica napus] Length = 356
810	2023810	1E-82 >sp P29525 OLEO_ARATH OLEOSIN >gi 282875 pir S22538 oleosin - Arabidopsis thaliana >gi 16405 emb CAA44225 (X62353) oleosin [Arabidopsis thaliana] >gi 4455257 emb CAB36756.1 (AL035523) oleosin, 18.5K [Arabidopsis thaliana] Length = 207
811	2023811	1E-108 >gi 4056502 (AC005896) 40S ribosomal protein S5 [Arabidopsis thaliana] Length = 207
812	2023812	1E-123 >gi 3319357 (AF077407) contains similarity to phosphoenolpyruvate synthase (ppsA) (GB:AE001056) [Arabidopsis thaliana] Length = 662
813	2023813	7E-55 >emb CAB06417 (Z84377) xylosidase [Aspergillus niger] Length = 804
814	2023814	3E-11 >gi 3548810 (AC005313) chloroplast nucleoid DNA binding protein [Arabidopsis thaliana] Length = 461
815	2023815	3E-33 >gi 3402683 (AC004697) patatin-like protein [Arabidopsis thaliana] Length = 499
816	2023816	6E-92 >sp P49209 RL9_ARATH 60S RIBOSOMAL PROTEIN L9 >gi 2129720 pir S71255 ribosomal protein L9 - Arabidopsis thaliana >gi 1107489 emb CAA63024 (X91958) 60S ribosomal protein L9 [Arabidopsis thaliana] Length = 195
817	2023817	1E-10 >emb CAB38212 (AL035601) protein [Arabidopsis thaliana] Length = 252
818	2023818	1E-130 >gi 2618688 (AC002510) esterase D [Arabidopsis thaliana] Length = 284
819	2023819	1E-171 >sp P46644 AAT3_ARATH ASPARTATE AMINOTRANSFERASE, CHLOROPLAST PRECURSOR (TRANSAMINASE A) >gi 693692 (U15034) aspartate aminotransferase [Arabidopsis thaliana] Length = 449
820	2023820	1E-17 >dbj BAA33206 (AB001888) zinc finger protein [Oryza sativa] Length = 407
821	2023821	Tyr Phospho Site(160-167)
822	2023822	1E-122 >gi 2388578 (AC000098) Similar to Mycobacterium RlpF (gb Z84395). ESTs gb T75785,gb R30580,gb T04698 come from this gene. [Arabidopsis thaliana] Length = 223
823	2023823	1E-129 >gb AAD25665.1 AC007020_7 (AC007020) ferritin protein [Arabidopsis thaliana] >gi 4588004 gb AAD25945.1 AF085279_18 (AF085279) hypothetical ferritin subunit [Arabidopsis thaliana] Length = 259
824	2023824	Zinc Finger C2h2(360-382)
825	2023825	2E-91 >gi 3688799 (AF057137) gamma tonoplast intrinsic protein 2 [Arabidopsis thaliana] Length = 253
826	2023826	Tyr Phospho Site(60-67)
827	2023827	6E-68 >sp P32110 GTX6_SOYBN PROBABLE GLUTATHIONE S-TRANSFERASE (HEAT SHOCK PROTEIN 26A) (G2-4) >gi 99912 pir A33654 heat shock protein 26A - soybean >gi 169981 (M20363) Gmhsp26-A [Glycine max] Length = 225
828	2023828	1E-101 >gb AAD39666.1 AC007591_31 (AC007591) Is a member of the PF 00903 glyoxalase family. ESTs gb T44721, gb T21844 and gb AA395404 come from this gene. [Arabidopsis thaliana] Length = 174
829	2023829	Rgd(1357-1359)
830	2023830	5E-90 >gb AAD30232.1 AC007202_14 (AC007202) Is a member of the PF 00171 aldehyde dehydrogenase family. ESTs gb T21534, gb N65241 and gb AA395614 come from this gene. [Arabidopsis thaliana] Length = 509
831	2023831	2E-20 >sp Q46036 BLC_CITFR OUTER MEMBRANE LIPOPROTEIN BLC

		PRECURSOR >gi 2121019 pir I40710 outer membrane lipoprotein - Citrobacter freundii >gi 717136 (U21727) lipocalin precursor [Citrobacter freundii] Length = 177
832	2023832	2E-89 >sp P30707 RL9_PEA 60S RIBOSOMAL PROTEIN L9 (GIBBERELLIN-REGULATED PROTEIN GA) >gi 100065 pir S19978 ribosomal protein L9 - garden pea >gi 20727 emb CAA46273 (X65155) GA [Pisum sativum] Length = 193
833	2023833	Tyr Phospho Site(896-903)
834	2023834	2E-87 >sp P42748 UBC4_ARATH UBIQUITIN-CONJUGATING ENZYME E2-21 KD 1 (UBIQUITIN-PROTEIN LIGASE 4) (UBIQUITIN CARRIER PROTEIN 4) >gi 431266 (L19354) ubiquitin conjugating enzyme [Arabidopsis thaliana] Length = 187
835	2023835	9E-83 >gi 1256424 (U51119) cysteine proteinase inhibitor [Brassica campestris] Length = 205
836	2023836	1E-119 >gb AAD50015.1 AC007651_10 (AC007651) glutathione transferase [Arabidopsis thaliana] Length = 221
837	2023837	Zinc Finger C2h2(1242-1265)
838	2023838	Tyr Phospho Site(88-96)
839	2023839	Pkc Phospho Site(31-33)
840	2023840	1E-180 >gi 3355490 (AC004218) dolichyl-phosphate beta-glucosyltransferase [Arabidopsis thaliana] Length = 336
841	2023841	1E-101 >gi 682728 (L40031) S-adenosyl-L-methionine:trans-caffeoyl-Coenzyme A 3-O-methyltransferase [Arabidopsis thaliana] Length = 212
842	2023842	3E-14 >gi 3293547 (AF072709) oxidoreductase [Streptomyces lividans] Length = 313
843	2023843	5E-25 >dbj BAA82843.1 (AB023651) miraculin homologue [Solanum melongena] Length = 160
844	2023844	1E-110 >sp P54888 P5C2_ARATH DELTA 1-PYRROLINE-5-CARBOXYLATE SYNTHETASE B (P5CS B) [INCLUDES: GLUTAMATE 5-KINASE (GAMMA-GLUTAMYL KINASE) (GK); GAMMA-GLUTAMYL PHOSPHATE REDUCTASE (GPR) (GLUTAMATE-5-SEMIALDEHYDE DEHYDROGENASE) (GLUTAMYL-GAMMA-SEMIALDE... >gi 887388 emb CAA60447 (X86778) pyrroline-5-carboxylate synthetase B [Arabidopsis thaliana] >gi 1669658 emb CAA70527 (Y09355) pyrroline-5-carboxylate synthetase [Arabidopsis thaliana] Length = 726
845	2023845	1E-138 >gi 1020155 (U26936) DNA-binding protein [Arabidopsis thaliana] Length = 236
846	2023846	4E-76 >emb CAB38956.1 (AL049171) pyrophosphate-dependent phosphofructo-1-kinase [Arabidopsis thaliana] Length = 500
847	2023847	1E-155 >gi 4185136 (AC005724) trehalose-6-phosphate synthase [Arabidopsis thaliana] Length = 862
848	2023848	1E-30 >gi 2642215 (AF030386) NOI protein [Arabidopsis thaliana] Length = 79
849	2023849	2E-59 >gi 2739044 (AF024651) polyphosphoinositide binding protein Ssh1p [Glycine max] Length = 324
850	2023850	2E-59 >sp P40602 APG_ARATH ANTER-SPECIFIC PROLINE-RICH PROTEIN APG PRECURSOR >gi 99694 pir S21961 proline-rich protein APG - Arabidopsis thaliana >gi 22599 emb CAA42925 (X60377) APG [Arabidopsis thaliana] Length = 534
851	2023851	Pkc Phospho Site(5-7)
852	2023852	1E-104 >gi 3395434 (AC004683) peroxidase [Arabidopsis thaliana] >gi 742248 prf 2009327B peroxidase [Arabidopsis thaliana] Length = 349
853	2023853	Tyr Phospho Site(1115-1122)
854	2023854	6E-40 >dbj BAA76393.1 (AB025187) cytochrome c oxidase subunit 6b-1 [Oryza sativa] Length = 169
855	2023855	Tyr Phospho Site(426-433)

856	2023856	6E-43 >pir S52995 arabinogalactan-like protein - loblolly pine >gi 607774 (U09556) arabinogalactan-like protein [Pinus taeda] Length = 264
857	2023857	3E-91 >sp P47997 G11A_ORYSA PROTEIN KINASE G11A >gi 100705 pir B30311 protein kinase C (EC 2.7.1.-) homolog - rice (fragment) >gi 169788 (J04556) G11A protein [Oryza sativa] Length = 531
858	2023858	3E-93) >gi 3927825 (AC005727) dTDP-glucose 4-6-dehydratase [Arabidopsis thaliana] Length = 343
859	2023859	1E-101 >gb AAD41971.1 AC006438_3 (AC006438) cold acclimation protein WCOR413 [Triticum aestivum] [Arabidopsis thaliana] Length = 197
860	2023860	1E-137 >emb CAB37533 (AL035538) glycine hydroxymethyltransferase like protein [Arabidopsis thaliana] Length = 517
861	2023861	1E-112) >gi 4056502 (AC005896) 40S ribosomal protein S5 [Arabidopsis thaliana] Length = 207
862	2023862	6E-98 >gi 4204274 (AC004146) ribulose biphosphate carboxylase, small subunit [Arabidopsis thaliana] Length = 180
863	2023863	4E-76 >pir S71286 oleosin isoform - Arabidopsis thaliana >gi 987014 emb CAA90877 (Z54164) oleosin [Arabidopsis thaliana] >gi 987016 emb CAA90878 (Z54165) oleosin [Arabidopsis thaliana] Length = 191
864	2023864	Pkc Phospho Site(42-44)
865	2023865	Tyr Phospho Site(974-982)
866	2023866	Tyr Phospho Site(355-362)
867	2023867	6E-35 >dbj BAA18248 (D90912) ferredoxin [Synechocystis sp.] Length = 122
868	2023868	Tyr Phospho Site(109-117)
869	2023869	Tyr Phospho Site(638-645)
870	2023870	5E-30 >emb CAB55502.1 (AJ131768) tyramine hydroxycinnamoyltransferase [Nicotiana tabacum] Length = 226
871	2023871	1E-131 >emb CAB45850.1 (AL080254) reticuline oxidase-like protein [Arabidopsis thaliana] Length = 539
872	2023872	9E-99) >emb CAB41123.1 (AL049657) argininosuccinate synthase-like protein [Arabidopsis thaliana] Length = 498
873	2023873	Tyr Phospho Site(1364-1370)
874	2023874	1E-108 >gb AAD32833.1 AC007659_15 (AC007659) mitochondrial elongation factor G [Arabidopsis thaliana] Length = 754
875	2023875	1E-66 >emb CAA65533 (X96758) clathrin coat assembly protein AP17 [Zea mays] Length = 132
876	2023876	3E-92 >sp Q43117 KPYA_RICCO PYRUVATE KINASE ISOZYME A, CHLOROPLAST PRECURSOR >gi 169703 (M64736) ATP:pyruvate phosphotransferase [Ricinus communis] Length = 583
877	2023877	4E-83 >emb CAB10235.1 (Z97336) auxin-responsive protein IAA1 [Arabidopsis thaliana] Length = 168
878	2023878	2E-33 >gi 3822225 (AF079183) RING-H2 finger protein RHG1a [Arabidopsis thaliana] Length = 190
879	2023879	1E-24 >gb AAD38289.1 AC007789_15 (AC007789) ABA induced plasma membrane protein [Oryza sativa] Length = 189
880	2023880	1E-105 >sp P10797 RBS3_ARATH RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN 2B PRECURSOR (RUBISCO SMALL SUBUNIT 2B) >gi 68061 pir RKMUB2 ribulose-bisphosphate carboxylase (EC 4.1.1.39) small chain B2 precursor - Arabidopsis thaliana >gi 16194 emb CAA32701 (X14564) ribulose biphosphate carboxylase [Arabidopsis thaliana] Length = 181
881	2023881	1E-139 >gi 3402678 (AC004697) adenylate kinase [Arabidopsis thaliana] Length = 295
882	2023882	Tyr Phospho Site(98-106)
883	2023883	5E-26 >gb AAD34267.1 AF084419_1 (AF084419) calmodulin mutant SYNCAM64A [synthetic construct] Length = 147

884	2023884	2E-15 >bbs 48073 13 kDa-B polypeptide of iron-sulfur protein fraction of NADH:ubiquinone oxidoreductase [cattle, heart, Peptide Mitochondrial Partial, 114 aa] Length = 114
885	2023885	Tyr Phospho Site(937-944)
886	2023886	4E-73 >gb AAD39281.1 AC007576_4 (AC007576) initiation factor 5A-4 [Arabidopsis thaliana] Length = 158
887	2023887	Pkc Phospho Site(69-71)
888	2023888	Tyr Phospho Site(100-106)
889	2023889	6E-74 >emb CAB38706.1 (AJ131464) nitrate transporter [Arabidopsis thaliana] Length = 567
890	2023890	Tyr Phospho Site(1268-1275)
891	2023891	Zinc Finger C2h2(755-775)
892	2023892	7E-81 >dbj BAA24074 (D89824) GTP-binding protein [Arabidopsis thaliana] Length = 210
893	2023893	2E-33 >gi 4164539 (AF079170) phloem protein [Cucurbita maxima] Length = 150
894	2023894	4E-15 >gi 2739366 (AC002505) SF16 like protein [Arabidopsis thaliana] Length = 516
895	2023895	Phospho Site(1301-1307)
896	2023896	1E-57 >emb CAA74052 (Y13724) Transcription factor [Arabidopsis thaliana] Length = 187
897	2023897	Tyr Phospho Site(768-775)
898	2023898	5E-38 >gi 3599491 (AF085149) aminotransferase [Capsicum chinense] Length = 459
899	2023899	Rgd(210-212)
900	2023900	Tyr Phospho Site(1201-1208)
901	2023901	1E-144 >pir S51697 oleoyl-[acyl-carrier-protein] hydrolase (EC 3.1.2.14) - Arabidopsis thaliana >gi 2129530 pir S69195 acyl-(acyl carrier protein) thioesterase (clone TE 1-1) - Arabidopsis thaliana >gi 634003 emb CAA85387 (Z36910) acyl-(acyl carrier protein) thioesterase [Arabidopsis thaliana] Length = 412
902	2023902	5E-79 >gi 2281629 (AF003095) AP2 domain containing protein RAP2.2 [Arabidopsis thaliana] Length = 246
903	2023903	5E-91 >sp Q39836 GBLP_SOYBN GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKE PROTEIN >gi 1256608 gb AAB05941.1 (U44850) G beta-like protein [Glycine max] Length = 325
904	2023904	7E-87 >gi 1872544 (U89014) early light-induced protein; ELIP [Arabidopsis thaliana] Length = 195
905	2023905	1E-108 >gi 507164 (U04818) PITSLRE alpha 2-4 [Homo sapiens] Length = 562
906	2023906	1E-121 >gi 3421082 (AF043523) 20S proteasome subunit PAD2 [Arabidopsis thaliana] Length = 250
907	2023907	6E-69 >sp P55964 KPYG_RICCO PYRUVATE KINASE ISOZYME G, CHLOROPLAST Length = 418
908	2023908	1E-108 >gi 3033400 (AC004238) Ser/Thr protein kinase [Arabidopsis thaliana] Length = 1257
909	2023909	1E-127 >gb AAD31337.1 AC007354_10 (AC007354) Strong similarity to gb Y09533 involved in starch metabolism from Solanum tuberosum and contains a PF01326 Pyruvate phosphate dikinase, PEP/pyruvate binding domain. EST gb N96757 comes from this gene. [... Length = 1358
910	2023910	Tyr Phospho Site(1347-1355)
911	2023911	Tyr Phospho Site(1324-1331)
912	2023912	Rgd(731-733)
913	2023913	5E-31 >gb AAD20708 (AC006300) glucose-induced repressor protein [Arabidopsis thaliana] Length = 628

914	2023914	Tyr_Phospho_Site(4-11)
915	2023915	3E-30 >emb CAB38807.1 (AL035678) nucellin-like protein [Arabidopsis thaliana] Length = 420
916	2023916	3E-50 >dbj BAA22813 (D26015) CND41, chloroplast nucleoid DNA binding protein [Nicotiana tabacum] Length = 502
917	2023917	5E-67 >gi 2281633 (AF003097) AP2 domain containing protein RAP2.4 [Arabidopsis thaliana] Length = 229
918	2023918	2E-98 RBS4_ARATH RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN SUBUNIT
919	2023919	Sugar_Transport_2(364-389)
920	2023920	Tyr_Phospho_Site(218-225)
921	2023921	3E-41 >emb CAB51834.1 (AJ243961) contains eukaryotic protein kinase domain PF 00069 [Oryza sativa] Length = 844
922	2023922	4E-28 >gb AAD28599.1 AF126742_1 (AF126742) bundle sheath defective protein 2 [Zea mays] Length = 129
923	2023923	2E-75 >gi 1408473 (U48939) actin depolymerizing factor 2 [Arabidopsis thaliana] Length = 137
924	2023924	1E-91 >dbj BAA20084.1 (AB003590) sulfate transporter [Arabidopsis thaliana] >gi 2114106 dbj BAA20085.1 (AB003591) sulfate transporter [Arabidopsis thaliana] Length = 677
925	2023925	5E-88 >gi 2317912 (U89959) cathepsin B-like cysteine proteinase [Arabidopsis thaliana] Length = 357
926	2023926	Tyr_Phospho_Site(591-597)
927	2023927	1E-110 >emb CAA16940.1 (AL021768) small GTP-binding protein-like [Arabidopsis thaliana] Length = 200
928	2023928	1E-112 >gb AAD28774.1 AF134127_1 (AF134127) Lhcb4.2 protein [Arabidopsis thaliana] Length = 287
929	2023929	4E-54 >emb CAB56149.1 (AJ242970) BTF3b-like factor [Arabidopsis thaliana] Length = 165
930	2023930	5E-21 >gb AAD46412.1 AF096262_1 (AF096262) ER6 protein [Lycopersicon esculentum] Length = 168
931	2023931	1E-105 >sp P10797 RBS3_ARATH RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN 2B PRECURSOR (RUBISCO SMALL SUBUNIT 2B) >gi 68061 pir I RK MUB2 ribulose-bisphosphate carboxylase (EC 4.1.1.39) small chain B2 precursor - Arabidopsis thaliana >gi 16194 emb CAA32701 (X14564) ribulose bisphosphate carboxylase [Arabidopsis thaliana] Length = 181
932	2023932	Tyr_Phospho_Site(1153-1159)
933	2023933	2E-82 >gi 3834310 (AC005679) Similar to Ubiquitin-conjugating enzyme E2-17 KD gb D83004 from Homo sapiens. ESTs gb T88233, gb Z24464, gb N37265, gb H36151, gb Z34711, gb AA040983, and gb T22122 come from this gene. [Arabidopsis thaliana] Length = 163
934	2023934	1E-104 >gb AAB51571.1 (U75193) germin-like protein [Arabidopsis thaliana] >gi 1755168 gb AAB51573.1 (U75195) germin-like protein [Arabidopsis thaliana] >gi 2239042 emb CAA73213 (Y12673) GLP3 protein [Arabidopsis thaliana] Length = 206
935	2023935	Tyr_Phospho_Site(1372-1379)
936	2023936	1E-106 >emb CAB41927.1 (AL049751) ribosomal protein L13a like protein [Arabidopsis thaliana] Length = 206
937	2023937	Pkc_Phospho_Site(51-53)
938	2023938	3E-79 >sp O65788 C7B2_ARATH CYTOCHROME P450 71B2 >gi 3164140 dbj BAA28537 (D78605) cytochrome P450 monooxygenase [Arabidopsis thaliana] Length = 502
939	2023939	Tyr_Phospho_Site(11-18)
940	2023940	Tyr_Phospho_Site(13-20)
941	2023941	6E-57 >pir S52578 protein-serine/threonine kinase NPK15 - common tobacco >gi 505146 dbj BAA06538 (D31737) protein-serine/threonine kinase

		[<i>Nicotiana tabacum</i>] Length = 422
942	2023942	8E-94) >gi 3337356 (AC004481) protein transport protein SEC61 alpha subunit [<i>Arabidopsis thaliana</i>] Length = 475
943	2023943	4E-38 >gi 2459440 (AC002332) receptor kinase [<i>Arabidopsis thaliana</i>] Length = 664
944	2023944	6E-14 >sp P80728 MAVI_CUCPE MAVICYANIN >gi 1836088 bbs 179249 mavicyanin=12.752 kda small blue copper-containing stellacyanin-like glycoprotein/type I cupredoxin [<i>Cucurbita pepo</i> =green zucchini, peelings, Peptide, 108 aa] Length = 108
945	2023945	5E-60 >gb AAD34695.1 AC006341_23 (AC006341) Similar to gb AJ224359 surfait locus protein 5 (surf5b) from <i>Homo sapiens</i> . [<i>Arabidopsis thaliana</i>] Length = 150
946	2023946	Tyr Phospho Site(257-264)
947	2023947	1E-78) >emb CAB10195.1 (Z97335) transport protein [<i>Arabidopsis thaliana</i>] Length = 769
948	2023948	1E-39 >gi 3386612 (AC004665) DNA-binding protein, dbp [<i>Arabidopsis thaliana</i>] Length = 190
949	2023949	Pkc Phospho Site(12-14)
950	2023950	Tyr Phospho Site(574-580)
951	2023951	1E-55 >pir S37101 ATAF1 protein - <i>Arabidopsis thaliana</i> (fragment) >gi 1345506 emb CAA52771 (X74755) ATAF1 [<i>Arabidopsis thaliana</i>] Length = 229
952	2023952	Pkc Phospho Site(45-47)
953	2023953	1E-125 >emb CAB38921.1 (AL035709) bZIP transcription factor-like protein [<i>Arabidopsis thaliana</i>] Length = 305
954	2023954	1E-93 >emb CAA72792 (Y12071) thylakoid lumen rotamase [<i>Spinacia oleracea</i>] Length = 449
955	2023955	7E-64) >gi 2708746 (AC003952) DnaJ-like chaperonin [<i>Arabidopsis thaliana</i>] Length = 160
956	2023956	9E-95 >pir S33612 isocitrate dehydrogenase - soybean Length = 451
957	2023957	1E-106 >sp O23515 RL15_ARATH 60S RIBOSOMAL PROTEIN L15 >gi 2245027 emb CAB10447.1 (Z97341) ribosomal protein [<i>Arabidopsis thaliana</i>] Length = 204
958	2023958	1E-63 >gb AAC28488.1 (AF079588) 1-aminocyclopropane-1-carboxylate oxidase [<i>Sorghum bicolor</i>] Length = 316
959	2023959	3E-58 >emb CAB36546.1 (AL035440) DNA binding protein [<i>Arabidopsis thaliana</i>] Length = 427
960	2023960	Tyr Phospho Site(190-196)
961	2023961	Tyr Phospho Site(818-825)
962	2023962	1E-131 >gi 2511725 (AF021937) catalase 1 [<i>Arabidopsis thaliana</i>] Length = 492
963	2023963	1E-19 >gi 1905887 (U92461) recombination factor GdRad54 [<i>Gallus gallus</i>] Length = 733
964	2023964	1E-103 >sp P46283 S17P_ARATH SEDOHEPTULOSE-1,7-BISPHOSPHATASE, CHLOROPLAST PRECURSOR (SEDOHEPTULOSE-BISPHOSPHATASE) (SBPASE) (SED(1,7)P2ASE) >gi 1076403 pir S51838 sedoheptulose-1,7-biphosphatase - <i>Arabidopsis thaliana</i> >gi 786
965	2023965	2E-17 >emb CAA99819.1 (Z75533) waek similarty with bacillus amyloliquefaciens permease IIBC (Swiss Prot accession number P41029); cDNA EST yk573h3.3 comes from this gene [<i>Caenorhabditis elegans</i>] Length = 378
966	2023966	8E-26 >pir S49463 chloroplast RNA binding protein - kidney bean >gi 558629 emb CAA57551 (X82030) chloroplast RNA binding protein [<i>Phaseolus vulgaris</i>] Length = 287
967	2023967	1E-44 >emb CAA55397 (X78820) casein kinase I [<i>Arabidopsis thaliana</i>] Length = 364

968	2023968	1E-105) >gb AAB51565.1 (U75187) germin-like protein [Arabidopsis thaliana] Length = 204
969	2023969	2E-96 >emb CAA65502 (X96727) isocitrate dehydrogenase (NAD+) [Nicotiana tabacum] Length = 364
970	2023970	Pkc Phospho Site(26-28)
971	2023971	4E-43 >gi 871780 (L43080) pEARL1 1 gene product [Arabidopsis thaliana] >gi 4725947 emb CAB41718.1 (AL049730) pEARL1 1 [Arabidopsis thaliana] Length = 168
972	2023972	2E-16 >sp P24805 TSJT_TOBAC STEM-SPECIFIC PROTEIN TSJT1 >gi 100383 pir S13551 stem-specific protein - common tobacco >gi 20037 emb CAA36525 (X52283) stem specific, weakly expressed in other organs [Nicotiana tabacum] Length = 149
973	2023973	1E-18 >gb AAD21041 (AF116237) pseudouridine synthase 1 [Mus musculus] Length = 393
974	2023974	Tyr Phospho Site(95-102)
975	2023975	1E-108) >prf 1804333B Gln synthetase [Arabidopsis thaliana] Length = 430
976	2023976	1E-116 >gi 2947070 (AC002521) Ser/Thr protein kinase [Arabidopsis thaliana] Length = 429
977	2023977	3E-15 >sp P74523 YE19_SYNY3 HYPOTHETICAL 17.7 KD PROTEIN SLR1419 >gi 1653717 dbj BAA18628 (D90916) hypothetical protein [Synechocystis sp.] Length = 159
978	2023978	7E-20 >gi 3033400 (AC004238) Ser/Thr protein kinase [Arabidopsis thaliana] Length = 1257
979	2023979	Tyr Phospho Site(28-35)
980	2023980	Pkc Phospho Site(16-18)
981	2023981	Rgd(231-233)
982	2023982	Pkc Phospho Site(16-18)
983	2023983	3E-24 >gi 2854070 (AF044914) histone deacetylase [Arabidopsis thaliana] Length = 305
984	2023984	1E-28 >gi 3157924 (AC002131) Contains homology to extensin-like protein gb D83227 from Populus nigra. ESTs gb H76425, gb T13883, gb T45348, gb H37743, gb AA042634, gb Z26960 and gb Z25951 come from this gene. There is a similar ORF on the opposite strand. [... >gi 4063707 (AF104327) extensin-like protein [Arabidopsis thaliana] Length = 137
985	2023985	Receptor Cytokines 1(1550-1562)
986	2023986	1E-113 >gi 3420055 (AC004680) cyclophilin [Arabidopsis thaliana] Length = 201
987	2023987	2E-27 >emb CAB45075.1 (AL078637) serine/threonine kinase-like protein [Arabidopsis thaliana] Length = 445
988	2023988	Zinc Finger C2h2(929-950)
989	2023989	1E-141 >pir S37495 peroxidase (EC 1.11.1.7) - Arabidopsis thaliana >gi 405611 emb CAA50677 (X71794) peroxidase [Arabidopsis thaliana] Length = 353
990	2023990	Tyr Phospho Site(1189-1197)
991	2023991	5E-92 >sp P28148 TF22_ARATH TRANSCRIPTION INITIATION FACTOR TFIID-2 (TATA-BOX FACTOR 2) (TATA SEQUENCE-BINDING PROTEIN 2) (TBP-2) >gi 99764 pir S10945 transcription initiation factor IID (clone At-1) - Arabidopsis thaliana >gi 16546 emb CAA38742 (X54995) transcription initiation factor II [Arabidopsis thaliana] >gi 4204264 (AC005223) 43453 [Arabidopsis thaliana] >gi 227073 prf 1613452A transcription initiation factor TFIID-1 [Arabidopsis thaliana] Length = 200
992	2023992	3E-16 >gi 3790581 (AF079179) RING-H2 finger protein RHB1a [Arabidopsis thaliana] Length = 190
993	2023993	1E-20 >sp Q28735 TM21_RABIT TRANSMEMBRANE PROTEIN TMP21

		PRECURSOR (INTEGRAL MEMBRANE PROTEIN P23) >gi 1370279 emb CAA66947 (X98303) transmembrane protein [Oryctolagus cuniculus] Length = 219
994	2023994	Tyr Phospho Site(112-119)
995	2023995	3E-11 >gb AAD35009.1 AF144391_1 (AF144391) thioredoxin-like 5 [Arabidopsis thaliana] Length = 185
996	2023996	Tyr Phospho Site(1372-1379)
997	2023997	7E-12 >sp P40389 UV22_SCHPO UV-INDUCED PROTEIN UVI22 >gi 629909 pir S47147 uvi22 protein - fission yeast (Schizosaccharomyces pombe) >gi 1076930 pir JC2442 UV inducible protein, UV122 - fission yeast (Schizosaccharomyces pombe) >gi 499199 emb CAA84069 (Z34299) uvi22 [Schizosaccharomyces pombe] >gi 3184086 emb CAA19342 (AL023781) uv-induced protein uvi22 [Schizosaccharomyces pombe] Length = 303
998	2023998	2E-28 >sp P30185 DH18_ARATH DEHYDRIN RAB18 >gi 282880 pir S28021 rab18 protein - Arabidopsis thaliana >gi 16451 emb CAA48178 (X68042) RAB18 [Arabidopsis thaliana] Length = 186
999	2023999	4E-93 >sp P42795 R111_ARATH 60S RIBOSOMAL PROTEIN L11A (L16A) >gi 624938 emb CAA57395 (X81799) ribosomal protein L16 [Arabidopsis thaliana] Length = 182